



Ana Luísa Dos Santos Magalhães
Licenciada em Ciências Biomédicas

Identificação de Novos Genes de Susceptibilidade para o Cancro do Cólon e Recto do Tipo X

Dissertação para obtenção do Grau de Mestre em
Genética Molecular e Biomedicina

Orientadora: Doutora Maria Cristina Mantas Albuquerque
Valeroso,
Investigadora, Unidade de Investigação em Patobiologia
Molecular (UIPM), Instituto Português de Oncologia de
Lisboa Francisco Gentil, EPE (IPOLFG)

[Identificação de Novos Genes de Susceptibilidade para o Cancro do Cólon e Recto do tipo X]

Copyright [Ana Luísa dos Santos Magalhães], FCT/UNL, UNL

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Em memória aos meus avós

“....Por muito longe, por muito difícil, por muito desesperadamente. O coração guarda o que se nos escapa das mãos. E durante o dia e durante a vida, quando não está lá quem se ama, não é ela que nos acompanha- é o nosso amor, o amor que se lhe tem. Não é para perceber. É sinal de amor puro não se perceber, amar e não se ter, querer e não guardar a esperança, doer sem ficar magoado, viver sozinho, tristes, mas mais acompanhado de quem vive feliz. Não se pode ceder. Não se pode resistir. A vida é uma coisa, o amor é outra...”

Miguel Esteves Cardoso

Agradecimentos

Em primeiro lugar, queria deixar aqui os meus agradecimentos às pessoas sem as quais a realização deste projecto não seria possível.

À Dr. Cristina Albuquerque, por me ter dado a oportunidade de fazer parte do seu grupo de trabalho, por acreditar sempre nas minhas capacidades e por toda a sua disponibilidade, simpatia, força e conhecimento que me transmitiu.

À Mestre Inês Francisco, um agradecimento especial não só pela sua ajuda no trabalho prático e na revisão da tese, mas também por todo o carinho, apoio e conhecimento que me deu ao longo deste ano.

À Dr. Patrícia Silva, ao Dr. Bruno Filipe e à Mestre Marlene Duarte, por toda a ajuda que me deram no laboratório e pela simpatia com que me receberam.

Às minhas colegas de laboratório Soraia Magnost, Carolina Pereira e Íris Guerreiro por todo o espírito de companheirismo e auxílio no laboratório.

À Doutora Branca Cavaco, Coordenadora da unidade de Investigação em Patobiologia Molecular, do Instituto Português de Oncologia, Lisboa, Francisco Gentil, E.P.E, por ter concedido a oportunidade de participar neste projecto.

Ao Serviço de Gastrenterologia, Clínica de Risco familiar e ao Serviço de Anatomia Patológica do Instituto Português de Oncologia, Lisboa, Francisco Gentil, E.P.E, pela disponibilização do material biológico e informação clínica dos doentes.

À Dr. Marta Pojo, investigadora da Unidade de Investigação em Patobiologia Molecular, do Instituto Português de Oncologia, Lisboa, Francisco Gentil, E.P.E, por toda a sua disponibilidade e ajuda neste projecto.

Às minhas colegas de casa Cátia Rodrigues, Fabiana Domingos e Denise Francisco por todo o suporte, paciência e pela amizade incomparável.

À minha irmã “adoptada” Inês Mesquita, e ao Miguel Casteleiro, pela coragem de aturarem o meu difícil feitio e por nunca me deixarem fraquejar.

Aos meus pais, o maior agradecimento de todos, por serem os melhores, e por todo o amor e suporte que me dão.

Resumo

O cancro do cólon e recto familiar do tipo X (FCCTX) descreve as famílias HNPCC que preenchem os CA mas não apresentam mutações germinais nos genes MMR e cujos tumores são microssatélites estáveis. Não são ainda conhecidas causas moleculares que expliquem a susceptibilidade para o desenvolvimento de cancro do cólon na grande maioria das famílias FCCTX. Assim, o presente projecto pretendeu identificar genes candidatos/variantes específicas em genes candidatos que possam estar envolvidos na susceptibilidade para o cancro do cólon e recto familiar do tipo X (FCCTX).

Foram incluídas neste estudo e foco principal das várias análises, 15 famílias FCCTX. Com base em resultados prévios do grupo e numa possível interacção com o gene *MSH6*, foi estudada a contribuição de alterações genéticas germinais no gene *FBXO11* para a susceptibilidade para o FCCTX, por sequenciação de Sanger, análise de *copy-number* e de expressão na linha germinal. Para as mesmas famílias foi efectuada uma caracterização a nível germinal de dois subgrupos de famílias FCCTX, correspondentes a duas entidades moleculares previamente identificadas pelo grupo (TSG⁺ e TSG⁻), através de sequenciação de nova geração utilizando um painel multigénico de 94 genes que conferem risco aumentado para cancro. Com base nos resultados desta análise foi seleccionado um sub-grupo de genes, que foram analisados para 78 indivíduos index adicionais de potenciais famílias FCCTX. Para uma família, mais informativa, do grupo TSG⁻, foi avaliado o potencial carácter patogénico de variantes genéticas específicas em genes candidatos para a susceptibilidade para o FCCTX, obtidas através de *whole exome sequencing* (WES), de modo a identificar mutações que possam predispor para esta condição hereditária. Com esta finalidade foi efectuada análise bioinformática e análise *in silico* para selecção de variantes candidatas seguida de análise de segregação com a doença na família em 9 de indivíduos afectados e 3 não afectados.

Os resultados obtidos da análise mutacional, de expressão e de *copy-number* não revelaram qualquer contribuição do gene *FBXO11* para o FCCTX. A análise do painel multigénico para as famílias FCCTX identificou mutações em genes que codificam para proteínas associadas a vias de reparação do DNA associadas à recombinação homóloga em 7/17 (47%), mas apenas em famílias com tumores com assinatura molecular TSG⁺. Mutações nestes genes foram também identificadas num subgrupo de potenciais famílias FCCTX (29/78, 37%), o que poderá vir a ser importante para o manejo clínico destas famílias. A análise das variantes provenientes da sequenciação do exoma numa família FCCTX, permitiu a selecção de 43 variantes, das quais, após completa análise *in silico* e de segregação, se seleccionaram 6 variantes nos genes *MTMR3*, *DUSP12*, *LGR6*, *SMG7*, *TAS1R1* e *NEK7*, como contribuindo possivelmente para o FCCTX. A identificação de mais do que uma variante genética em algumas famílias parece estar de acordo com um modelo do FCCTX como uma doença poligénica.

Palavras-Chave: FCCTX, Genes de Susceptibilidade, *FBXO11*, Sequenciação de Exoma

Abstract

Familial colorectal cancer type X (FCCTX) defines families that fulfil the Amsterdam criteria, but without any germline mutation in the DNA mismatch repair genes and whose tumours do not present microsatellite instability. However, the hereditary genetic basis for this disease is not defined yet. Thus, the present project aimed to identify candidate genes/specific variants in candidate genes that may be involved in susceptibility to FCCTX.

We included in this study 15 FCCTX families that were the main focus of the various analyses. Based on previous results and in its possible interaction with the *MSH6* gene, the contribution of germline genetic alterations in the *FBXO11* gene for susceptibility to FCCTX was studied by Sanger sequencing, copy-number analysis and gene expression at the germline level. For the same families, a germline characterization of two subgroups of FCCTX families, corresponding to two molecular entities previously identified (TSG + and TSG-), was carried out through new generation sequencing using a multigene panel of 94 genes that confer increased risk for cancer. Based on the results of this analysis, a subgroup of genes was selected, which were analyzed for 78 additional index subjects from potential FCCTX families. For a more informative family from the TSG- group, the potential pathogenicity of specific genetic variants in candidate genes for susceptibility to FCCTX, obtained through whole exome sequencing (WES), was evaluated in order to identify mutations that may predispose for this hereditary condition. For this purpose, bioinformatic and *in silico* analysis were performed for selection of candidate variants, followed by segregation analysis with the disease in the family in 9 affected and 3 unaffected relatives.

Results obtained from mutational, expression and copy-number analysis did not reveal any contribution of the *FBXO11* gene to FCCTX. Multigene panel analysis for FCCTX families identified mutations in genes that code for proteins associated with homologous recombination-DNA repair pathways in 7/17 (47%), but only in families with TSG + molecular signature tumours. Mutations in these genes were also identified in a subgroup of potential FCCTX families (29/78, 37%), which may be important for the clinical management of these families. Analysis of the variants obtained from the exome sequencing in an FCCTX family allowed the selection of 43 variants, which, after complete *in silico* and segregation analysis, enabled the selection of 6 variants in the *MTMR3*, *DUSP12*, *LGR6*, *SMG7*, *TASIR1* and *NEK7* genes as possibly contributing to FCCTX. The identification of more than one genetic variant in some families appears to be in agreement with a model of FCCTX as a polygenic disease.

Keywords: FCCTX, Susceptibility Genes, *FBXO11*, Exome Sequencing

Índice

1. Introdução	1
1.1. Cancro do cólon e recto	1
1.2. Mecanismos subjacentes à carcinogénese do CCR	1
1.2.1. A sequência adenoma-carcinoma	1
1.2.2. Vias de instabilidade envolvidas no CCR	3
1.3 Síndromes hereditárias associadas ao CCR	4
1.3.1 Síndromes associadas a polipose	5
1.3.2. Cancro do cólon e recto hereditário não associado a polipose	7
1.3.2.1. Síndrome de Lynch	8
1.3.2.2. Cancro do cólon e recto familiar do tipo X	8
1.4. Envolvimento do gene <i>FBXO11</i> como candidato para a susceptibilidade para o FCCTX	10
1.5. Caracterização molecular de um grupo de famílias FCCTX – trabalho prévio do grupo de Gastreterologia da UIPM	13
1.5.1. Estudo das vias de instabilidade genómica em tumores de famílias FCCTX	13
1.5.2. Caracterização a nível germinal, das famílias FCCTX	14
1.5.2.1. Caracterização a nível germinal, das famílias FCCTX (subgrupos TSG ⁺ e TSG ⁻) através da análise de um painel multigénico por <i>Next Generation Sequencing</i>	14
1.5.2.2. Identificação de novos <i>loci</i> de susceptibilidade para o FCCTX por análise de <i>linkage</i> em todo o genoma	15
1.5.2.3. Identificação de novas variantes para a susceptibilidade para o FCCTX através de <i>whole- exome sequencing</i>	15
2. Objectivos	21
3. Material e Métodos	23
3.1. Material Biológico	23
3.2. Métodos	24
3.2.1. Extração de ácidos nucleicos	24
3.2.1.1. Extração de DNA genómico e RNA a partir de sangue periférico	24
3.2.1.2. Extração de DNA e RNA a partir de tecido tumoral e mucosa normal incluído em parafina	24
3.2.1.3. Extração de DNA a partir de tecido congelado	25
3.2.1.4. Extração de RNA a partir de células da linha de carcinoma do cólon <i>SW480</i>	25
3.2.2. Reacção em cadeia da polimerase (PCR)	25
3.2.2.1. Optimizações das condições de PCR	26
3.2.3. Análise do gene <i>FBXO11</i> como candidato para a suscetibilidade para o FCCTX	27
3.2.3.1. Análise mutacional do gene <i>FBXO11</i>	27
3.2.3.1.1. Amplificação do DNA genómico por PCR	27
3.2.3.1.2. Controlo da eficiência dos produtos amplificados por PCR em gel de agarose 2% (p/v) ..	28
3.2.3.1.3. Purificação enzimática dos produtos amplificados por PCR	28
3.2.3.1.4. Sequenciação automática	29
3.2.3.1.4.1. Reacção de Sequenciação	29
3.2.3.1.4.2. Precipitação e purificação do DNA após reacção de sequenciação	30
3.2.3.1.4.3. Preparação das amostras para electroforese capilar no sequenciador automático utilizando o <i>software Sequencing Analysis</i>	30
3.2.3.1.4.4. Análise de resultados	31
3.2.3.2. <i>Real-Time</i> PCR	31
3.2.3.2.1. Optimização das condições da reacção de qPCR e cálculo da sua eficiência	32

3.2.3.2.2. Análise de <i>copy number</i> para o gene <i>FBXO11</i>	33
3.2.3.2.3. Análise de expressão do gene <i>FBXO11</i>	34
3.2.3.2.3.1. Reacção de Transcrição Reversa (RT-PCR) para síntese de cDNA.....	34
3.2.4. Caracterização a nível germinal, das famílias FCCTX (subgrupos TSG ⁺ e TSG ⁻) através da análise de um painel multigénico por <i>Next Generation Sequencing</i>	35
3.2.4.1. Análise mutacional utilizando um painel multigénico e selecção das variantes possivelmente patogénicas	36
3.2.4.1.1. Desenho de <i>primers</i> específicos para a análise das várias variantes presentes em estudo	37
3.2.4.1.2. Confirmação das variantes seleccionadas, por sequenciação de Sanger	37
3.2.5. Identificação de genes candidatos para a susceptibilidade para o cancro do cólon e recto familiar do tipo X (FCCTX)	38
3.2.5.1. Selecção de variantes identificadas através da análise de <i>whole exome sequencing</i> na família L56.....	38
3.2.5.1.1. Análise de segregação das variantes seleccionadas com a doença, na família L56.....	39
3.2.5.1.1.1. Análise mutacional somática da variante p.Arg645Trp do gene <i>MTMR3</i> na família L56	40
3.2.5.1.1.2. Análise de expressão do gene <i>MTMR3</i> em tumores por <i>Real-Time</i> PCR	41
4. Resultados	43
4.1. Envolvimento do gene <i>FBXO11</i> na susceptibilidade para FCCTX.....	43
4.1.1. Análise Mutacional	43
Legenda: As frequências alélicas foram obtidas da base de dados <i>Ensembl</i> , excepto as assinaladas* que foram obtidas da base de dados <i>ExAC</i> (<i>Exome Aggregation Consortium</i>). Para as variantes que apresentavam uma baixa frequência na população procedeu-se a uma caracterização adicional recorrendo aos <i>softwares</i> <i>VEP</i> e/ou <i>Mutation Taster</i> , dependendo do tipo de mutação.....	
4.1.2. Análise de <i>copy number</i> para o gene <i>FBXO11</i>	44
4.1.3. Análise de Expressão do gene <i>FBXO11</i>	45
.....	46
4.2 Caracterização a nível germinal, das famílias FCCTX (subgrupos TSG ⁺ e TSG ⁻) através da análise de um painel multigénico por <i>Next Generation Sequencing</i>	46
4.3 Identificação de novos genes candidatos para a susceptibilidade para o FCCTX através de sequenciação do exoma.....	50
4.3.1 Selecção de variantes identificadas através da sequenciação do exoma	50
4.3.2 Estudo de segregação das variantes seleccionadas, identificadas através da sequenciação do exoma.....	53
4.3.3. Análise somática da variante p.Arg645Trp identificada no gene <i>MTMR3</i>	62
4.3.4 Análise de expressão do gene <i>MTMR3</i> e <i>RAC1</i> no tumor de um indivíduo portador da mutação germinal neste gene	62
5. Discussão	64
5.1. Análise do envolvimento do gene <i>FBXO11</i> na susceptibilidade para FCCTX	64
5.2. Caracterização a nível germinal, das famílias FCCTX (subgrupos TSG ⁺ e TSG ⁻) através da análise de um painel multigénico por <i>Next Generation Sequencing</i>	64
5.3 Identificação de novos genes candidatos para a susceptibilidade para o FCCTX por sequenciação do exoma.....	67
6. Conclusão.....	76
7. Referências	78
8 Anexos	86

Índice de Figuras

Figura 1.1. Esquema representativo da via de sinalização WNT.	2
Figura 1.2. Sequência adenoma-carcinoma e os vários genes e mecanismos envolvidos (.....	3
Figura 1.3. Localização do gene <i>FBXO11</i> no cromossoma 2 e a representação dos domínios da proteína codificada por este gene.....	10
Figura 1.4. Representação esquemática do envolvimento do gene <i>FBXO11</i> em alguns mecanismos associados ao cancro.	11
Figura 1.5. Representação esquemática dos genes MSH6 e FBXO11 e a sua sobreposição..	12
Figura 1.6. Resumo das principais diferenças moleculares identificadas nas duas entidades moleculares distintas (TSG ⁺ e TSG ⁻) descritas nas famílias FCCTX.	13
Figura 1.7. Árvore geneológica da família L56.	18
Figura 3.1. Gráfico de amplificação para uma amostra por qPCR e representação das suas respectivas fases. (adaptado de:Arya et al.,2005)	31
Figura 3.2. Critérios de exclusão utilizados para a selecção das variantes para estudo de segregação com a doença na família L56.	38
Figura 4.1. Excertos de electroforetogramas representativos de 3 das 7 variantes identificadas.	43
Figura 4.2. Resultados obtidos através da análise de <i>copy number</i> para o gene <i>FBXO11</i> normalizada contra o controlo <i>Ch12 probe</i> , representado pela linha a azul.	45
Figura 4.3. Resultados obtidos em relação à expressão relativa do gene <i>FBXO11</i>	46
Figura 4.4. Os indivíduos assinalados a cores foram seleccionados para análise bioinformática, no que diz respeito à análise 1.	54
Figura 4.5. Os indivíduos assinalados a cores foram seleccionados para análise bioinformática, no que diz respeito à análise 3.	56
Figura 4.6. Os indivíduos assinalados a cores foram seleccionados para análise bioinformática, no que diz respeito à análise 4.	58
Figura 4.7. Os indivíduos assinalados a cores foram seleccionados para análise bioinformática, no que diz respeito à análise 5.	60
Figura 4.8. Os indivíduos assinalados a cores foram seleccionados para análise bioinformática, no que diz respeito à análise bioinformática previamente realizada ao estudo.	61
Figura 4.9. Representação cromossómica dos vários genes onde foram identificadas as variantes	61
Figura 4.10. Análise de expressão para os genes <i>MTMR3</i> , <i>RAC1</i> e <i>RAC1b</i>	63
Figura 5.1. Mecanismo sugerido pelo presente estudo, envolvido na reparação de erros no DNA causados por stress genotóxico em FCCTX.....	66
Figura 5.2. (A) Representação da desfosforilação do fosfatidilinositol PtdIns3P e do PtdIns(3,5)P2 pela MTMR3; (B) Representação dos domínios proteicos da MTMR3 e a localização da variante identificada no contexto da análise de exoma.	68

Figura 5.3. Representação dos domínios proteicos com os vários sítios de fosforilação (vermelho) da proteína e a variante identificada pela análise de exoma	69
Figura 5.4. representação esquemática do modelo de função da proteína NEK7 no ciclo celular.	70
Figura 5.5. Possível mecanismo subjacente à conversão de GTP em cGMP pela proteína NPR2.	71
Figura 5.6. Representação dos domínios proteicos do GC-B e identificação da variante p.Val882Ile na respectiva região do domínio CYC da proteína NPR2.	71
Figura 5.7. Representação da possível interacção do LGR6 com a RSPO e a sua participação na via <i>WNT</i>	72
Figura 5.8. Representação da interacção das proteínas DUSP12 e HSP70 e o seu possível envolvimento na inibição da apoptose.	73
Figura 5.9. Acção dos receptores TAS1R1/TAS1R3 na presença de aminoácidos e a sua consequente activação da proteína MTORC1 de forma a inibir a autofagia	75

Índice de Tabelas

Tabela 1.1. Critérios de Amesterdão I e II e critérios de Bethesda	7
Tabela 1.2. Descrição dos <i>outputs</i> obtidos após a análise bioinformática e respetivo número de variantes total obtido	19
Tabela 1.3 Descrição da estratégia utilizada nas cinco análises bioinformáticas efectuadas a partir dos resultados obtidos por WES, para os indivíduos seleccionados da família L56 e respetivo número de variantes obtido... ..	19
Tabela 3.1. Critérios usados para a selecção das variantes como possivelmente patogénicas, utilizando os <i>softwares</i> VEP (<i>SIFT</i> e <i>Polyphen-2</i>) e <i>Mutation Taster</i>	36
Tabela 3.2. Caracterização das amostras de DNA dos indivíduos da família L56 incluídas na análise mutacional somática do gene <i>MTMR3</i>	40
Tabela 4.1. Caracterização das variantes obtidas na análise mutacional dos 22 exões codificantes do gene <i>FBXO11</i> , para as 15 famílias FCCTX e para os doentes cujo tumor apresentava ausência de expressão da proteína MSH6.	44
Tabela 4.2. Descrição e caracterização das mutações germinais potencialmente patogénicas identificadas nos 15 indivíduos index de famílias FCCTX através da análise de um painel multigénico (Trusight Cancer).	47
Tabela 4.3. Descrição dos tumores apresentados nas famílias FCCTX estudadas e classificação clínica e molecular dos tumores dos 15 indivíduos index de famílias FCCTX, estudados através do painel multigénico.	48
Tabela 4.4. Resultados da análise mutacional dos genes de reparação do DNA do painel em doentes de famílias HNPCC e em doentes que preenchem os CB.	49
Tabela 4.5. Resumo do número de variantes iniciais obtidas pela análise bioinformática, após a aplicação dos critérios de selecção e análise <i>in silico</i> nas diferentes análises e <i>outputs</i>	50
Tabela 4.6. Caracterização das variantes seleccionadas após a aplicação dos critérios de selecção e critérios mais restritos em termos de frequência das mutações e de análise <i>in silico</i> (ponto 3.2.5.1).	51
Tabela 4.7. Resultados da análise de segregação das variantes seleccionadas a partir dos resultados de <i>output</i> da análise 1.	53
Tabela 4.8. Resultados da análise de segregação das variantes seleccionadas a partir dos resultados de <i>output</i> da análise 2.	54
Tabela 4.9. Resultados da análise de segregação das variantes seleccionadas a partir dos resultados de <i>output</i> da análise 3.	55
Tabela 4.10. Resultados da análise de segregação das variantes seleccionadas a partir dos resultados de <i>output</i> da análise 4. Na tabela estão descritos os genes onde se identificaram as variantes. Para	

cada indivíduo foi realizada uma breve caracterização clínica, incluindo o número de carcinomas e de pólipos.	57
Tabela 4.11. Resultados da análise de segregação das variantes seleccionadas a partir dos resultados de <i>output</i> da análise 5.....	58
Tabela 4.12. Resultados obtidos para a análise somática da variante p.Arg645Trp do gene <i>MTMR3</i> ..	62

Lista de acrónimos

% (p/v) – percentagem expressa em peso por volume

APAF- Polipose adenomatosa familiar atenuada

APC- *Adenomatous polyposis coli*

ASS- Adenoma serreado sésil

AST- Adenoma serreado tradicional

BCL2- *B-Cell CLL/Lymphoma 2*

BCL6- *B-Cell CLL/Lymphoma 6*

BER- *Base excision repair*

BMP- *Bone morphogenetic protein signalling*

BMP4- *Bone morphogenetic protein 4*

BMPRI1- *Bone morphogenetic protein receptor type 1A*

CA- Critérios de Amesterdão

CA-I- Critérios de Amesterdão I

CA-II- Critérios de Amesterdão II

CB- Critérios de Bethesda

CCR- Cancro do cólon e recto

cDNA-Dna complementar

CHRPE- Hipertrofia congénita do epitélio pigmentar da retina

CIN- Instabilidade Cromossómica

CIPM- Fenótipo metilador das ilhas CpG

CK1- *Casein kinase 1*

CUL1-*Cullin 1*

DCC- *DCC Netrin 1 Receptor*

ddNTP's – Didesoxirribonucleótidos

DNA – Ácido desoxirribonucleico

dNTP's – Desoxirribonucleótidos

EMT- Transição epitélio-mesenquima

EPCAM- *Epithelial Cell Adhesion Molecule*

FBXO11- *F-box only protein 11*

FCCTX- Cancro do cólon e recto familiar do tipo X

GALNT12- *Polypeptide N-*

acetylgalactosaminyltransferase 12

GNAS-GNAS complex Locus

GSK3- *Glycogen syntahse kinase 3*

HABP4-*Hyaluronan binding protein 4*

HNPCC- Cancro do cólon e recto hereditário não associado a polipose

KRAS- *V-Ki-ras2 Kristen Rat Sarcoma Viral Oncogene Homolog*

LOH- Perda de heterozigotia

MAPK- *Mitogen-activated protein kinase*

MDM2- *MDM2 proto-oncogene*

MDMX- actual MDM4, p53 regulator

MGMT-O-6-Methylguanine-DNA

Methyltransferase

MLH1- MutL homolog 1

MMR- Genes de reparação de erros no DNA do tipo *mismatch*

MS-MLPA-*Methylation specific – multiplex ligation-dependent probe amplification*

MSH2- *MutS homolog 2*

MSH6- MutS homolog 6

MSI- Instabilidade de Microsatélites

MSI-H- Instabilidade de Microsatélites de alto grau

MSI-L- Instabilidade de Microsatélites de baixo grau

MSS- Microsaélite estável

MUTYH- *Muty DNA glycosylase*; homólogo do *mutY (E. coli)*

NGS- Sequenciação de nova geração

PAF- Polipose adenomatosa familiar

PAM- Polipose adenomatosa associada ao gene *MUTYH*

pb – pares de base

PH- Pólipo Hiperplásico

PJS- Síndrome Peutz-Jeghers

PMS2- PMS1 homolog 2

PODL1- DNA polymerase delta 1, catalytic subunit

POLE- DNA polymerase epsilon, catalytic subunit

PPAF- Polymerase proofreading-associated polyposis

PRTM9-Protein arginine N-methyltransferase 9

PS-Polipose serreada

PTEN- Phosphatase and tensin homolog

RBX1-Ring Box 1

rpm – Rotações por minuto

RPS20- Ribossomal Protein S20

SEMA4A- Semaphorin 4

SKP1-S-Phase Kinase-Associated Protein 1

SL- Síndrome de Lynch

SMAD2- Mothers against decapentaplegic homolog 2

SMAD4- Mothers against decapentaplegic homolog 4

SNAIL- Snail Family Zinc Finger SKP1- S-phase kinase-associated protein 1

STK11-Serine/Threonine Kinase 11

TE – tampão Tris-EDTA

TGF- β – Transforming growth factor beta

Tm – Temperatura de melting

TP53- Tumor protein P53

TSG⁻ - Tumor supressor gene negative

TSG⁺ - Tumor supressor gene positive

VEP – *Variant Effect Predictor*

VIT-1- Vitiligo-associated protein 1

WES- *Whole exome sequencing*

WGS- *Whole genome sequencing*

WHO- Organização Mundial de Saúde

1. Introdução

1.1. Cancro do cólon e recto

O cancro do cólon e recto (CCR) é considerado mundialmente o quarto cancro mais mortal, a seguir ao cancro do pulmão, estômago e fígado (Ong *et al.*, 2014). Em Portugal, o CCR é o segundo com maior incidência, a seguir ao cancro da mama nas mulheres e ao cancro da próstata nos homens, e a segunda causa de morte devida a cancro (Brandão e Lage, 2015). O CCR afecta igualmente ambos os sexos, mas apresenta uma distribuição geográfica assimétrica, com a incidência a revelar-se superior nos países desenvolvidos (Watson e Collins, 2011). As causas subjacentes ao CCR são complexas e heterogêneas. O risco para CCR depende de factores ambientais, nomeadamente uma dieta enriquecida em carne vermelha, consumo excessivo de álcool ou sedentarismo, mas também de factores genéticos (Roper e Hung, 2013). Aproximadamente 70% dos casos de CCR são esporádicos enquanto os restantes 30% estão associados a uma componente familiar (Gimena-Garcia *et al.*, 2015). Destes, apenas 5% estão relacionados com síndromes hereditárias bem caracterizadas clinicamente e molecularmente, enquanto os restantes 25% têm causa molecular desconhecida podendo estar associados a genes de baixa penetrância (Jasperson *et al.*, 2010).

O desenvolvimento tumoral é desencadeado por uma acumulação de alterações genéticas que levam à desregulação de processos celulares essenciais como a proliferação, diferenciação celular e apoptose (Watson e Collins, 2011). É assim essencial compreender os eventos moleculares subjacentes ao CCR para uma melhor prevenção da doença, diagnóstico e planeamento de estratégias de tratamento (Roper e Hung, 2013).

1.2. Mecanismos subjacentes à carcinogénese do CCR

1.2.1. A sequência adenoma-carcinoma

Em 1990, Fearon e Vogelstein propuseram um modelo para a tumorigénese coloretal, denominado por sequência adenoma-carcinoma que descreve o processo necessário para o desenvolvimento de CCR. Neste modelo, a progressão tumoral ocorre através de uma série de eventos genéticos, e consequentemente histológicos, com início na transformação do epitélio cólico normal num adenoma e posteriormente num carcinoma (Pino e Chung, 2010) (Roper e Hung, 2013).

Segundo este modelo, a inactivação do gene *APC* (*adenomatous polyposis coli*), é essencial para a iniciação tumoral, sendo este o primeiro evento genético a ocorrer. O gene *APC* participa na via de

sinalização *WNT* (figura 1.1.) e é responsável pela regulação da localização da β -catenina. A proteína APC faz parte de um complexo proteico (juntamente com a AXIN2, GSK3 (*glycogen synthase kinase 3*) e CK1 (*casein kinase 1*)) encarregue de fosforilar a β -catenina para posterior degradação no proteossoma. A inactivação do APC permite a acumulação da β -catenina no citoplasma, a qual é posteriormente translocada para o núcleo, onde executa o seu papel de factor de transcrição na activação de genes associados à proliferação e diferenciação celular. A activação destes genes é responsável pela alteração da homeostasia no epitélio cólico, devido à acumulação de células em proliferação, não diferenciadas, na cripta, o que origina a formação de um adenoma (Armaghany et al., 2012) (Basu et al., 2016) (Schepers e Clevers, 2012).

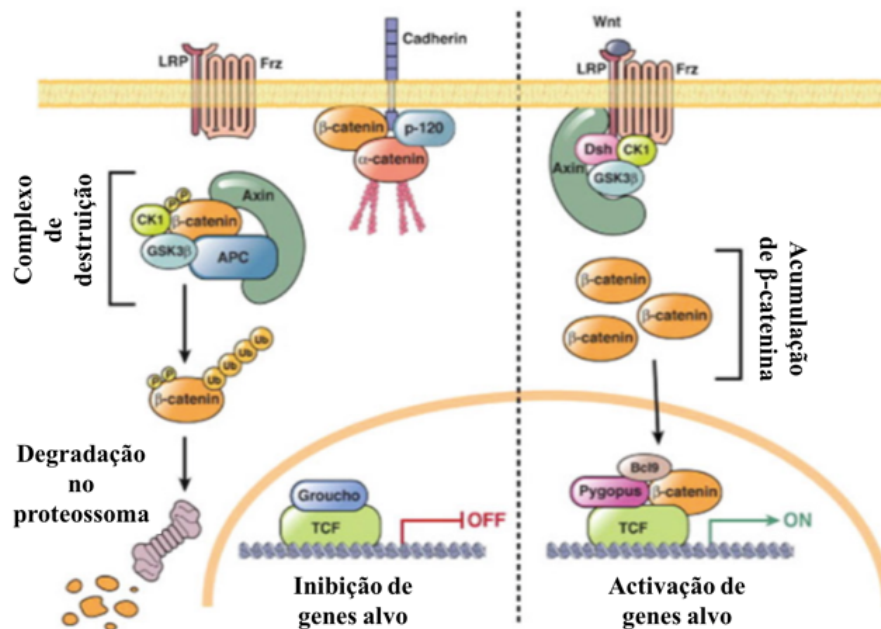


Figura 1.1. Esquema representativo da via de sinalização WNT. À esquerda: a via encontra-se na forma inactiva com consequente degradação da β -catenina no proteossoma; À direita: a via encontra-se na forma activa, ocorrendo acumulação de β -catenina no citoplasma, a qual é translocada para o núcleo e promove a activação de genes alvo. (adaptado de: Pino e Chung, 2010)

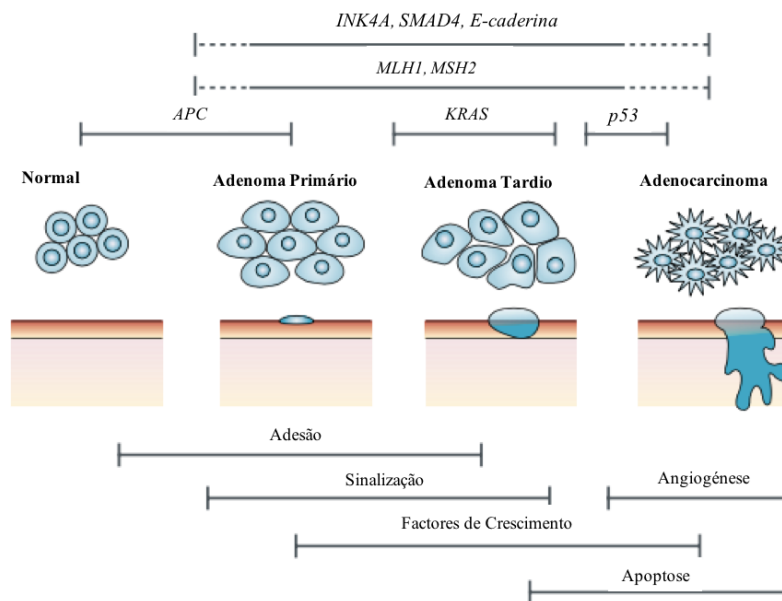


Figura 1.2. Sequência adenoma-carcinoma e os vários genes e mecanismos envolvidos (adaptado de: Kerr, 2003)

Associado à progressão do adenoma, ocorre a activação do proto-oncogene *KRAS* (*Kirsten Rat Sarcoma Viral Oncogene Homolog*), o qual faz parte da via MAPK (MAP kinase) que é essencial para a regulação da proliferação celular, diferenciação e apoptose (figura 1.2). Em seguida, estão descritas alterações no gene *TP53* (*tumor protein p53*), um gene supressor de tumor envolvido nas principais vias de regulação celular, como a apoptose, regulação do ciclo celular, angiogénese, diferenciação e migração celular. A perda de função da proteína p53 ocorre em cerca de 50-75% dos CCR, o que sugere que a inactivação bi-alélica (mutações e/ou perda de heterozigotia (LOH)) deste gene tem um papel crucial na transição de adenoma para carcinoma. Em estadios avançados podem ainda ocorrer outras alterações, sendo as mais frequentes a LOH em 18q, onde se localizam genes associados à via TGF- β (*transforming growth factor beta*) como o *SMAD4* (*SMAD Family Member 4*), o *SMAD2* (*SMAD Family Member 2*) e o gene *DCC* (*DCC netrin 1 receptor*) (Armaghany et al., 2012) (Roper e Hung, 2013).

1.2.2. Vias de instabilidade envolvidas no CCR

Para o desenvolvimento do CCR são necessários múltiplos eventos genéticos, sendo que a instabilidade genómica é actualmente reconhecida como uma característica celular essencial que acompanha a aquisição destas mutações. Na base da iniciação e progressão tumoral de CCR estão descritas três vias de instabilidade genómica, que diferem a nível genético e consequentemente nas manifestações clínicas e patológicas: a via da instabilidade cromossómica (CIN), a via da instabilidade

de microssatélites (MSI) e a via do fenótipo metilador das ilhas CpG (CIPM) (Pino e Chung, 2010) (Colussi et al., 2013).

A via CIN é a mais frequente, encontrando-se associada a cerca de 65-70% dos CCR esporádicos, os quais exibem um elevado número de perdas/ganhos da totalidade ou porções dos cromossomas, resultando em aneuploidia e frequente perda de heterozigotia (Roper e Hung, 2013). No geral, os tumores que seguem a via CIN estão associados a um pior prognóstico, comparativamente com os que seguem a via MSI (Colussi et al., 2013).

A MSI é detectada em cerca de 15% dos casos de CCR e caracteriza-se pela presença de erros em regiões repetitivas, abundantes no genoma, denominadas de microssatélites. Durante o processo de replicação do DNA, estas regiões são especialmente propícias a erros de *slippage* pela DNA polimerase, os quais são corrigidos pelo sistema de reparação de DNA do tipo *mismatch* (MMR). Na presença de um sistema deficiente, estes erros não são corrigidos resultando na chamada instabilidade de microssatélites, que se traduz pela presença de uma sequência de microssatélites com um tamanho diferente, na célula tumoral quando comparada com a célula normal. A inactivação do sistema MMR resulta de mutações nos genes *MSH2* (*MutS Homolog 2*), *MLH1* (*MutL Homolog 1*), *MSH6* (*MutS Homolog 6*) e *PMS2* (*PMS1 Homolog 2*), que quando ocorrem a nível germinal são responsáveis pela síndrome de Lynch (representa 3% dos casos de CCR com MSI) ou por hipermetilação do promotor do gene *MLH1* (associado à maioria dos CCR esporádicos que apresentam MSI) (Boland e Goel, 2010) (Church et al., 2016).

A CIMP apresenta-se como a terceira via de instabilidade genómica associada à progressão tumoral no CCR. Esta consiste na metilação aberrante (adição de grupos metil) das sequências dinucleotídicas CpG que se encontram, normalmente, nas regiões promotoras e/ou nos primeiros exões/intrões dos genes. Este evento epigenético leva ao silenciamento transcricional, afectando a expressão de vários genes envolvidos em processos celulares essenciais como a regulação do ciclo celular, apoptose, diferenciação e proliferação celular. Cerca de 20% - 30% dos CCR apresentam um fenótipo CIMP, sendo as características clínicas similares às associadas aos CCR que apresentam MSI. De notar que, contrariamente ao que acontece nas outras vias de instabilidade genómica, na CIMP a lesão precursora é um pólipó serrado e não um pólipó adenomatoso. (Church et al., 2016) (Ashktorab e Brim, 2014).

1.3 Síndromes hereditárias associadas ao CCR

As síndromes hereditárias de CCR dividem-se segundo a sua etiologia genética e manifestações clínicas, em dois grandes grupos: síndromes associadas a polipose e cancro do cólon e recto hereditário não associado a polipose (HNPCC).

1.3.1 Síndromes associadas a polipose

Das síndromes associadas a polipose fazem parte a polipose adenomatosa familiar (PAF), a polipose adenomatosa familiar atenuada (APAF), a polipose adenomatosa associada ao gene *MUTYH* (PAM), a *polymerase proofreading associated polyposis* (PPAP), a polipose serreada (PS) e a polipose hamartomatosa.

A polipose adenomatosa familiar (PAF) é uma condição hereditária autossômica dominante causada por mutações germinais no gene *APC*. Esta síndrome define-se pela presença de mais de cem adenomas por todo o cólon (frequentemente centenas a milhares), os quais, se não forem removidos conduzem a um risco aumentado de CCR que se aproxima de 100% na 4ª década de vida. Adicionalmente, os indivíduos com PAF podem também apresentar manifestações extracólicas como: tumores duodenais, pólipos gástricos, tumores desmóides, hipertrofia congénita do epitélio pigmentar da retina (CHRPE), osteomas e carcinoma papilar da tiróide. O desenvolvimento tumoral está associado à via CIN.

A polipose adenomatosa familiar atenuada (APAF) é uma condição menos severa que a PAF, também causada por mutações germinais no gene *APC*. Esta caracteriza-se por um menor número de adenomas (inferior a 100), diagnóstico numa idade mais tardia mas também com elevado risco de desenvolvimento de CCR, e elevada frequência de algumas das manifestações extracólicas associadas à PAF, como os tumores duodenais. Devido ao número de pólipos variável que a caracteriza (10-100), a APAF por vezes pode ser confundida com a PAM ou mesmo com casos esporádicos (Hawkins e Wise, 2016) (Jasperson et al., 2010) (Tomlinson, 2015).

A polipose adenomatosa associada ao gene *MUTYH* (PAM) tem uma apresentação fenotípica muito semelhante à APAF mas apresenta uma transmissão autossômica recessiva. Esta síndrome define-se molecularmente pela presença de mutações germinais bi-alélicas neste gene. O gene *MUTYH* (*mutY Dna glycosylase*) codifica para uma enzima que faz parte do sistema de reparação do DNA por excisão de bases (BER) envolvido na reparação de danos provocados por espécies reactivas de oxigénio. Funcionalmente, esta enzima ajuda a prevenir as transversões G:C para T:A causadas pelo *stress* oxidativo. A idade de diagnóstico da PAM varia entre os 40 a 60 anos, com os doentes a apresentarem um número variável de pólipos (geralmente inferior a 100). O risco de desenvolvimento de CCR é elevado, de cerca de 80%, em portadores bi-alélicos de mutações no *MUTYH*. Alguns estudos têm documentado um risco aumentado de desenvolvimento de CCR em portadores mono-alélicos, quando comparado com o risco de CCR esporádico na população em geral (Hawkins e Wise, 2016) (Tomlinson, 2015) (Jasperson et al., 2010), no entanto não é completamente consensual.

Recentemente com o avanço da sequenciação de nova geração (NGS), identificaram-se mutações germinais do tipo *missense* na região que codifica para os domínios de exonuclease dos genes *POLE* (*DNA polymerase epsilon, catalytic subunit*) e *POLD1* (*DNA polymerase delta 1, catalytic subunit*) em casos de CCR familiar. Estes domínios de exonuclease são importantes para a função de

proofreading na correcção de erros do tipo *mismatch* durante a replicação do DNA, sendo que estas enzimas intervêm na reparação dos erros antes do sistema MMR. Assim, foi descrita uma nova síndrome hereditária, denominada por *Polymerase proofreading-associated polyposis* (PPAP) com transmissão autossómica dominante, podendo apresentar algumas manifestações extracólicas e/ou múltiplos adenomas, sendo que, aparentemente os tumores não apresentam instabilidade de microssatélites (Tomlinson, 2015).

A polipose serreada (PS) é uma condição hereditária rara, podendo apresentar uma forma de transmissão autossómica dominante ou recessiva, e descrita por um fenótipo heterogéneo caracterizado pela presença de múltiplos pólipos serreados. Os pólipos apresentam assim uma arquitectura serreada nos quais se incluem: pólipos hiperplásicos (PH) (sem displasia ao contrário dos adenomas), adenomas serreados sésseis (ASS) (que podem ou não apresentar displasia) e adenomas serreados tradicionais (AST) que apresentam displasia (Huber et al., 2013) (Silva et al., 2016). Os critérios clínicos estabelecidos pela Organização Mundial de Saúde (WHO) para o diagnóstico da PS têm como base o número, o tamanho e localização dos pólipos serreados no cólon. A idade de diagnóstico média é de cerca de 51 anos, sendo que estes indivíduos apresentam um risco aumentado (cerca de 25-40%) para o desenvolvimento de CCR quando comparado com a população em geral. Na PS o risco para o desenvolvimento de outros tumores ou outras manifestações extracólicas é baixo. As causas genéticas para o desenvolvimento desta síndrome não são ainda conhecidas (Huber et al., 2013).

Por fim, a polipose hamartomatosa define-se pela presença de pólipos hamartomatosos normalmente de baixo potencial maligno e divide-se em síndrome Peutz-Jeghers (PJS) e polipose juvenil, ambas de transmissão autossómica dominante. A PJS está associada a mutações germinais no gene supressor de tumor *STK11* (*Serine/Threonine Kinase 11*). Este gene codifica para uma proteína que apresenta várias funções, como a regulação do ciclo celular, mediação da apoptose e polaridade celular. Mais de 95% dos indivíduos com PJS apresentam pigmentação cutânea nos lábios e mucosa bucal, uma característica fenotípica desta síndrome, sendo que os carcinomas no tracto gastrointestinal são frequentes (2-12%) (Huber et al., 2013). A polipose juvenil é definida pela presença de 5 ou mais pólipos juvenis no cólon ou associada a história familiar, independentemente do número de pólipos. Cerca de 35-60% dos casos ocorrem devido a mutações germinais no gene *SMAD4* e 15% no gene *BMPRI1A* (*Bone Morphogenetic Protein Receptor Type 1A*). Ambos os genes codificam para componentes da via de sinalização BMP (*bone morphogenetic protein signalling*) que é importante no processo de diferenciação das células da cripta e que está associada à via do TGF- β . As mutações no gene *SMAD4* estão associadas a uma doença mais severa no tracto gastrointestinal superior. O risco de desenvolvimento de CCR na polipose juvenil é cerca de 30-50 % (Hawkins e Wise, 2016) (Tomlinson, 2015).

1.3.2. Cancro do cólon e recto hereditário não associado a polipose

O cancro do cólon e recto hereditário não associado a polipose (HNPCC) corresponde a 1-5% dos casos de CCR, apresenta uma transmissão autossómica dominante e divide-se em síndrome de Lynch (SL) e cancro do cólon e recto familiar do tipo X (FCCTX) (Umar et al., 2004) (Lindor et al., 2005).

As famílias HNPCC são avaliadas clinicamente pelos critérios de Amesterdão I (CA-I) e II (CA-II) (definidos em 1990 e em 1999, respectivamente) que incluem a história familiar de CCR e um espectro de tumores extra-cólicos característicos desta síndrome (tabela 1.1). As famílias que não preenchem estes critérios são avaliados segundo os critérios de Bethesda (CB), de modo a identificar novos casos de SL (tabela 1.1.) (Boland, 2005). Caso haja preenchimento de apenas um CB é efectuada pesquisa de instabilidade de microssatélites no tumor através de dois marcadores de microssatélites de mononucleótidos BAT26 (2p21) e BAT25 (4q12) e três marcadores de dinucleótidos D2S123 (2p16), D5S346 (5q22-23) e D17S250 (17q12). O tumor é classificado como: MSI-Low (MSI-L) caso apresente instabilidade de microssatélites apenas num marcador; MSI-High (MSI-H) caso apresente instabilidade de microssatélites em dois ou mais marcadores ou microssatélite estável (MSS) quando não apresenta instabilidade de microssatélites em nenhum dos marcadores (Umar et al., 2004).

Tabela 1.1. Critérios de Amesterdão I e II e critérios de Bethesda. (adaptado de: Umar et al.,2004)

Critérios de Amesterdão I
<ol style="list-style-type: none">1. Pelo menos três familiares com CCR, sendo um deles familiar em 1º grau dos outros dois;2. Pelo menos duas gerações sucessivas afectadas;3. Pelo menos um dos tumores diagnosticado antes dos 50 anos;4. Exclusão de polipose adenomatosa familiar;
Critérios de Amesterdão II
<ol style="list-style-type: none">1. Pelo menos três familiares com carcinoma do cólon e recto (CCR) ou outro tumor associado a síndrome de Lynch (endométrio, intestino delgado, úter e pélvis renal) histologicamente confirmados, sendo um deles familiar em primeiro grau dos outros dois;2. Pelo menos duas gerações sucessivas afectadas;3. Pelo menos um dos tumores diagnosticado antes dos 50 anos;4. Exclusão de polipose adenomatosa familiar;
Critérios de Bethesda
<ol style="list-style-type: none">1. CCR diagnosticado antes dos 50 anos;2. Presença de CCR síncronos ou metacrónicos ou outro cancro associado a síndrome de Lynch, independentemente da idade;3. CCR diagnosticado antes dos 60 anos com histologia típica de síndrome de Lynch: presença de reacção de <i>Chron-like</i>, células em anel de sinete, invasão linfocitária e tumores com produção de muco;4. CCR diagnosticado e ≥ 1 familiar de 1º grau com cancro associado a síndrome de Lynch com pelo menos 1 dos cancros diagnosticado antes dos 50 anos5. CCR diagnosticado e ≥ 2 familiar de 1º grau com cancro associado a síndrome de Lynch, independentemente da idade;

1.3.2.1. Síndrome de Lynch

A síndrome de Lynch (SL) tem uma transmissão autossómica dominante e é caracterizada geneticamente por mutações germinais em genes envolvidos no sistema de reparação MMR, do qual fazem parte os genes *MSH2*, *MLH1*, *MSH6* e *PMS2*. Na maioria dos casos de SL verificam-se mutações no gene *MLH1* e *MSH2* (90%), enquanto mutações no gene *MSH6* correspondem aproximadamente a 6% dos casos e para o gene *PMS2* os casos são menos frequentes (Burt, 2007). Em algumas famílias com ausência de mutações germinais nos genes MMR, foram detectadas grandes deleções na região 3' do gene *EpCAM* (*Epithelial Cell Adhesion Molecule*), que se encontra a montante do *MSH2*, e que abrangiam também a região promotora deste, resultando num transcrito de fusão *EpCAM-MSH2*. Este é um novo mecanismo de inactivação germinal do *MSH2*, sendo a forma preferencial de inactivação somática a hipermetilação da região promotora (Ligtenberg et al., 2009).

O diagnóstico de SL é geralmente efectuado numa idade jovem (45 anos), o risco de desenvolvimento de CCR é elevado (50 a 80%) com os tumores a localizarem-se preferencialmente no cólon proximal e existe ainda uma elevada predisposição para o desenvolvimento de tumores extra-cólicos (como endométrio, intestino delgado, ureter e pélvis renal). Os CCR associados a esta síndrome têm características histológicas específicas como células em anel de sinete, presença da reacção de *Chron-like*, infiltrado linfocitário e tumores com produção de muco (Hawkins e Wise, 2016).

A nível clínico, avalia-se se as famílias preenchem os CA (tabela 1.1), para posterior pesquisa de mutações germinais nos genes MMR e caso se verifique a detecção de uma mutação patogénica, estas são então classificadas com SL (Jasperson et al., 2010) (Boland, 2005). Por outro lado, caso não preencham os CA é feita uma avaliação segundo os CB (tabela 1.1), sendo apenas necessário o preenchimento de um dos critérios para se proceder à pesquisa de instabilidade de microssatélites nos tumores como descrito em 1.3.2. (Umar et al., 2004). Cerca de 90% dos tumores de SL apresentam MSI, sendo esta uma característica da deficiência do sistema MMR (Jasperson et al., 2010). Caso os tumores sejam classificados como MSI-H, procede-se à análise mutacional dos genes MMR de modo a identificar novos casos de SL (Hawkins e Wise, 2016).

1.3.2.2. Cancro do cólon e recto familiar do tipo X

O cancro do cólon e recto familiar do tipo X (FCCTX) consiste nas famílias HNPCC que preenchem os CA mas não apresentam mutações germinais nos genes MMR e cujos tumores são microssatélite estáveis (Lindor et al., 2005). O FCCTX é uma condição autossómica dominante que em comparação com a SL apresenta um risco mais baixo para desenvolvimento de CCR, é diagnosticado numa idade mais tardia (50 a 60 anos) e sem risco aumentado para o desenvolvimento

de tumores extra-cólicos. Os tumores localizam-se predominantemente no cólon distal, sem produção de muco e infiltrado linfocitário (Hawkins e Wise, 2016) (Shiovitiz et al., 2014).

A causa molecular associada à susceptibilidade para o FCCTX ainda não foi identificada, contudo, têm sido realizados vários estudos com o intuito de clarificar os mecanismos implícitos a esta síndrome. Têm sido efectuados vários estudos de *whole-genome sequencing* (WGS) de forma a identificar novos *loci* de susceptibilidade, tendo sido já identificados alguns genes candidatos, como por exemplo, o *BMP4* (*bone morphogenetic protein 4*) em 8q23.3, o *HABP4* (*hyaluronan binding protein 4*) e o *GALNT12* (*polypeptide N-acetylgalactosaminyltransferase 12*), ambos no cromossoma 9. No entanto, até ao momento, os genes de susceptibilidade identificados são insuficientes para explicar por si só o FCCTX, e pensa-se que a combinação de mutações de baixo e moderado risco possam contribuir para a agregação familiar (Dominguez-Valentin et al., 2015).

Para além da identificação de possíveis *loci* de susceptibilidade, alguns genes/mutações foram, pontualmente, associados ao FCCTX. Num estudo de 2011, Nieminen e colaboradores, seleccionaram 18 famílias que preenchiam os CA-I ou CB sem mutação germinal nos genes MMR, e efectuaram uma análise de *genome-wide linkage* na família mais informativa (CA-I). Desta análise foi seleccionada a região do marcador D10S1686 (10q23), que contém entre outros, o gene *BMPRIA*. Este foi sequenciado, tendo sido identificada uma mutação patogénica (p.Glu188) que segregava com a doença na família. A análise foi posteriormente alargada às outras famílias, tendo sido identificada uma mutação de *splicing* (c.68-10_68+14del) que leva ao *skipping* do exão 2 (Nieminen et al., 2011). Na continuação deste estudo, foi efectuada uma análise de sequenciação de exoma numa outra família e identificaram uma mutação do tipo *frameshift* (p.Val50SerfsX23) no gene *RPS20* (*Ribossomal Protein S20*; 8q12.1) que segregava com a doença na família. Este gene foi estudado para as restantes famílias mas não foi identificada mais nenhuma mutação (Nieminen et al., 2014). A proteína RPS20 encontra-se associado à regulação da p53, e consequentemente à tumorigénese, uma vez que se liga e regula a MDM2 (*MDM2 proto-oncogene*) e a MDMX (actual *MDM4*, *p53 regulator*), importantes reguladores negativos da p53 (Daftuar et al., 2013) (Kim et al., 2014). Noutra família FCCTX foi identificada, por *whole exome sequencing* (WES), uma mutação do tipo *missense* (p.Val78Met) no gene *SEMA4A* (*Semaphorin 4A*), a qual revelou estar associada ao aumento significativo da activação das vias de sinalização MAPK/Erk e PI3K/Akt e consequente proliferação celular, na linha celular de cancro do cólon HCT116 (Schulz et al., 2014).

A dificuldade em identificar genes implicados no FCCTX levanta a hipótese que esta síndrome não seja monogénica, mas antes poligénica, resultando da interacção de vários genes de baixa penetrância (Stoffel e Kastrinos, 2014). Assim, apesar dos vários avanços efectuados para a caracterização do FCCTX, são ainda necessários estudos adicionais, no sentido da identificação da(s) causa(s) genética(s) associada(s) a esta síndrome.

1.4. Envolvimento do gene *FBXO11* como candidato para a susceptibilidade para o FCCTX

O gene *FBXO11* (*F-box only protein 11*) também conhecido por *PRTM9* (*Protein arginine N-methyltransferase 9*) e *VIT-1* (*Vitiligo-associated protein 1*) situa-se na região cromossômica 2p16.3 (figura 1.3), é constituído por 22 exões e tem um tamanho molecular de 3,8kb (Genecards: <http://www.genecards.org/>) (NCBI: <http://www.ncbi.nlm.nih.gov/>).

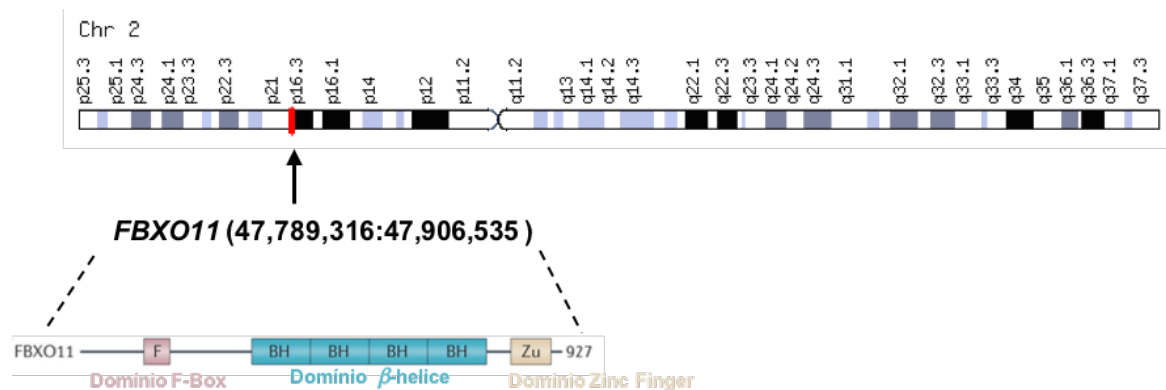


Figura 1.3. Localização do gene *FBXO11* no cromossoma 2 e a representação dos domínios da proteína codificada por este gene (adaptado de:(Wang et al., 2014) (Genecards: <http://www.genecards.org/>))

A proteína *FBXO11* faz parte de um subgrupo (*FBXO*) da família das proteínas *F-Box* e é um elemento do complexo proteico *Skp1 Cullin1 F-box* (*SCF*), responsável pela marcação de proteínas para degradação no proteossoma, por ubiquitinação (Wang et al., 2014) .

O complexo *SCF* é constituído pelas proteínas *SKP1* (*S-phase kinase-associated protein 1*), *RBX1* (*Ring Box 1*), *CUL1* (*cullin 1*) e uma proteína *F-Box* responsável pelo reconhecimento de um substrato específico. Estas proteínas reconhecem os seus substratos através de *dragons*, que são pequenos motivos específicos que se encontram em cada substrato e são responsáveis pela selecção das proteínas para degradação (Wang et al., 2014). O *FBXO11* tem sido referenciado como um potencial supressor tumoral e as suas interacções com outros genes influenciam vários processos celulares importantes para o desenvolvimento de cancro, nomeadamente, na proliferação e migração celular, conforme resume a figura 1.4 (Randle e Laman, 2016) (Wang et al., 2014) (Skaar et al, 2013).

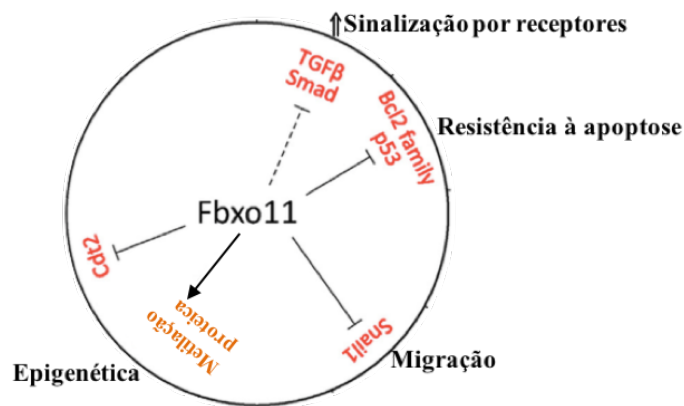


Figura 1.4. Representação esquemática do envolvimento do gene *FBXO11* em alguns mecanismos associados ao cancro: a inibição marcada a vermelho é mediada por ubiquitinação enquanto a laranja a inibição é independente deste processo (adaptado de: Randle e Laman, 2016).

Abida e colaboradores demonstraram que o complexo SCF-FBXO11 participa na inactivação da proteína p53 por conjugação, *in vivo* e *in vitro*, de uma proteína semelhante à ubiquitina (*Nedd8*) aos seus resíduos Lys-320 e Lys-321. A inibição do *FBXO11* leva a um aumento dos níveis de p21 (actual CDKN1A, *cyclin dependent kinase inhibitor 1A*), um alvo transcricional do *p53*, indicando assim que o *FBXO11* inibe a função do *p53* sem afectar a sua estabilidade (Abida et al., 2007). Adicionalmente, o FBXO11 interage também com a proteína CDT2 (*Denticleless E3 Ubiquitin Protein Ligase Homolog*) promovendo a sua degradação no proteosoma, o que leva à regulação do ciclo celular, em parte pela estabilização das proteínas p21 e SET8 (actual KMT5A, *lysine methyltransferase 5A*) importantes para a resposta das células epiteliais à estimulação induzida pelo TGF- β e para a migração celular (Heo et al., 2016).

Num estudo descrito por Duan e colaboradores, observou-se que a ubiquitinação da proteína BCL6 (B-Cell CLL/Lymphoma 6) também é mediada pelo complexo SCF-FBXO11. O oncogene *BCL6* controla a transcrição de uma variedade de genes envolvidos no desenvolvimento, diferenciação e activação dos linfócitos B. Mutações no gene *FBXO11*, que levem à perda da função da proteína, permitem a acumulação do *BCL6* o que pode conduzir ao desenvolvimento de DLBCL (*diffuse large B-cell lymphoma*) (Duan et al., 2012).

A transição epitélio-mesenquima (EMT) é um processo essencial para as células tumorais ganharem mobilidade e capacidades invasivas. A perda de função da proteína CDH1 (E-caderina) é um dos processos iniciais por detrás da EMT e a sua expressão pode ser reprimida pelo factor de transcrição *SNAIL* (*Snail Family Zinc Finger*). Recentemente o FBXO11 foi descrito como possível inibidor da metastização, e consequentemente potencial regulador da EMT, pois reconhece e marca para degradação a proteína SNAIL após a sua fosforilação pela proteína PKD1 (*Polycystin 1, Transient Receptor Potential Channel Interacting*) (Zheng et al., 2014).

Morrissy e colaboradores, identificaram a existência de uma interação do gene *FBXO11* com o gene *MSH6*, através de um mecanismo de expressão *antisense*. Este mecanismo é um exemplo de regulação entre genes que estão sobrepostos numa determinada região e com orientação oposta, como é o caso do *FBXO11* e do *MSH6* (figura 1.5). Este mecanismo foi confirmado pelo *splicing* de três sondas localizadas em dois exões do gene *MSH6*: exão 4 (2481163 e 2481164) e exão 5 (2481169). Quando os níveis de expressão do *FBXO11* são elevados, as regiões contendo as sondas 2481164 e 2481169 são excluídas durante o *splicing* enquanto para a sonda 2481163 (localizada no início do exão 4) observa-se a sua presença em isoformas do *MSH6* co-expressas com o gene *FBXO11*. Os resultados obtidos neste estudo sugerem que o *FBXO11* pode regular a expressão do *MSH6* por *splicing* alternativo (Morrissy et al., 2011).

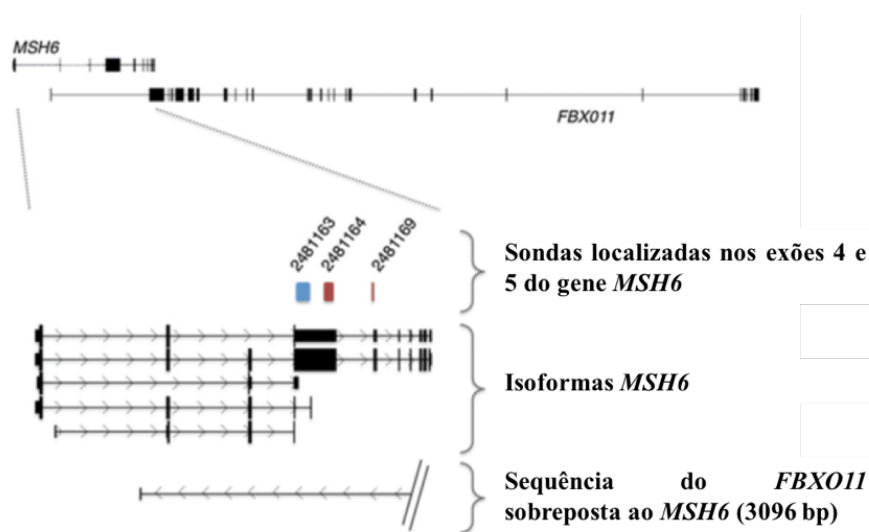


Figura 1.5. Representação esquemática dos genes *MSH6* e *FBXO11* e a sua sobreposição. Os quadrados a preto representam os exões enquanto as linhas entre os mesmos representam sequência não codificante (intrões); as sondas representadas a vermelho indicam uma correlação negativa com a elevada expressão do *FBXO11* enquanto que a sonda azul apresenta uma correlação positiva; (adaptado de: Morrissy et al., 2011).

Num estudo anterior, efectuado pelo grupo de Gastreenterologia da UIPM (Saramago,2014), foi avaliada a variação de *copy number* e a metilação por MS-MLPA (*Methylation specific – multiplex ligation-dependent probe amplification*), para os genes MMR, em 22 tumores FCCTX. Nestes, foi observada uma frequência elevada (55%) de deleções na região promotora do gene *MSH6*. Visto que o *FBXO11* e o *MSH6* estão sobrepostos nas regiões C-terminal, respectivas, e que as famílias em estudo não apresentam mutações germinais nos genes MMR, colocou-se a hipótese da ocorrência de um evento germinal no gene *FBXO11* e de um segundo evento (somático) corresponder às deleções observadas no gene *MSH6*. Para testar esta hipótese foi iniciada uma análise mutacional germinal para os 22 exões do gene *FBXO11*, a qual ainda qual não foi concluída (Duarte, 2015). De modo a compreender o envolvimento deste gene na susceptibilidade para FCCTX é fundamental finalizar esta análise mutacional e avançar com outros estudos adicionais como a análise de expressão do gene *FBXO11*.

1.5.1. Estudo das vias de instabilidade genómica em tumores de famílias FCCTX

De modo a contribuir para a caracterização do FCCTX, o grupo de Gastrenterologia da UIPM, tem vindo a desenvolver vários projectos nesta área. No trabalho realizado por Francisco et al., 2011 avaliou-se clinicamente e molecularmente um grupo homogéneo de famílias FCCTX registadas na Clínica de Risco Familiar do cancro do cólon e recto do IPOLFG, definidas pelos CA, ausência de mutações nos genes MMR e presença de tumores MSS. Este estudo foi efectuado em tumores de 15 famílias FCCTX, com o objectivo de avaliar o envolvimento da via CIN nos tumores associados a esta patologia. Para tal, foi avaliada a presença de LOH nos genes *APC*, *TP53*, *SMAD4* e *DCC*; análise de mutações somáticas nos genes *APC* e *KRAS* e estudo de metilação no promotor dos genes MMR e *MGMT* (*O*-6-Methylguanine-DNA-Methyltransferase). A análise de LOH permitiu observar dois grupos distintos de tumores entre as famílias FCCTX: 72% dos casos apresentavam perda de LOH nos genes supressores de tumor (TSG⁺: *tumor suppressor gene positive*) enquanto os restantes 28% não apresentavam esta característica (TSG⁻: *tumor suppressor gene negative*). Estes resultados indicam que a via CIN está associada ao desenvolvimento de tumores num sub-grupo de famílias FCCTX (TSG⁺). Os tumores deste sub-grupo apresentavam, com maior frequência mutações nos genes *APC* e *KRAS*, assim como metilação na região promotora dos genes MMR e/ou *MGMT*. Assim, estes resultados permitiram identificar a presença de duas entidades moleculares distintas entre as famílias FCCTX (Francisco et al., 2011).

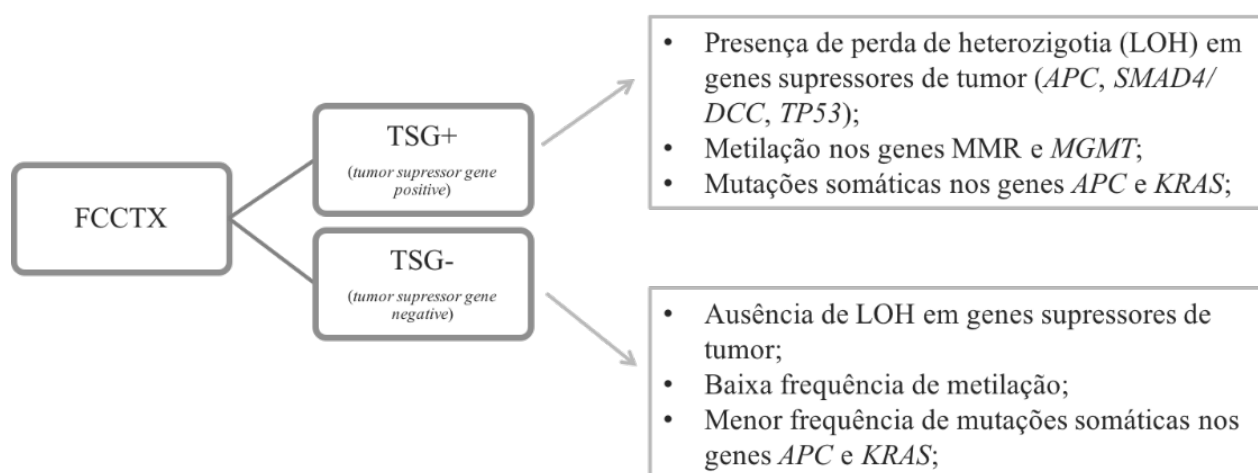


Figura 1.6. Resumo das principais diferenças moleculares identificadas nas duas entidades moleculares distintas (TSG⁺ e TSG⁻) descritas nas famílias FCCTX.

1.5.2. Caracterização a nível germinal, das famílias FCCTX

Considerando, no estudo anterior, a elevada prevalência de eventos somáticos observados no gene *APC* e do elevado número de adenomas nas famílias FCCTX, os autores colocaram a hipótese da contribuição de mutações germinais do tipo *missense* neste gene para estas famílias. Assim, a região do gene *APC* que abrange os três domínios de ligação à β -catenina (primeira metade do exão 15) foi analisada para os 15 indivíduos index, tendo sido identificada num indivíduo, de uma família TSG⁺ (L84), a mutação *missense* p.Glu1317Gln. A análise de segregação posteriormente efectuada, revelou que a mutação segregava com a doença na família. Esta mutação foi considerada como provável causa da doença nesta família, mas com as devidas precauções, associadas ao risco de desenvolvimento de CCR, devido à penetrância variável deste tipo de mutação, a qual pode ser modulada pelo ambiente e dieta (Francisco et al., 2011).

Recentemente, noutro trabalho realizado no grupo de Gastreenterologia, que avaliou o papel de mutações nos genes *POLE* e *POLD1* nas famílias FCCTX, foi identificada uma mutação possivelmente patogénica do tipo *missense* no gene *POLD1* (p.Pro300Leu) num dos indivíduos index das famílias FCCTX TSG⁺ (L37). Nesta família não foi possível realizar estudo de segregação da mutação com doença (Duarte, 2015).

1.5.2.1. Caracterização a nível germinal, das famílias FCCTX (subgrupos TSG⁺ e TSG⁻) através da análise de um painel multigénico por *Next Generation Sequencing*

Na continuação destes estudos, e de modo a caracterizar a nível germinal as 15 famílias FCCTX, nomeadamente os sub-grupos TSG⁺ e TSG⁻, foi efectuada sequenciação de nova geração, recorrendo ao painel comercial *Trusight Cancer (Illumina)*, que contém 94 genes associados a risco aumentado para cancro (anexo F). Como as mutações anteriormente identificadas nas famílias L84 e L37, não justificam por si só o fenótipo observado nas famílias, estas foram também incluídas neste estudo. A análise de NGS foi realizada pelo grupo de gastreenterologia da UIPM, na plataforma *MiSeq (Illumina)*. Este painel contém os genes MMR, pelo que se confirmou a ausência de mutações germinais nos genes *MLH1*, *MSH2* e *MSH6* (anteriormente analisados por sequenciação de Sanger) assim como noutros genes MMR, que não tinham sido avaliados anteriormente (*PMS2* e *PMS1*). Uma vez que mutações no gene *MUTYH* já foram reportadas em alguns casos de FCCTX (Woods et al., 2010), este gene foi avaliado nesta análise, não se tendo identificado nenhuma mutação bi-alélica nestas famílias. Assim, para completar a caracterização a nível germinal dos grupos TSG⁺ e TSG⁻, realizou-se a análise nos restantes genes do painel com o objectivo de identificar a possível causa molecular associada ao CCR nestas famílias, pelo que a análise dos resultados obtidos pela sequenciação deste painel de genes poderá contribuir para esclarecer a causa molecular da doença em algumas destas famílias.

1.5.2.2. Identificação de novos *loci* de susceptibilidade para o FCCTX por análise de *linkage* em todo o genoma

De modo a identificar novos *loci* de susceptibilidade para FCCTX, o grupo de gastroenterologia do IPOLFG, E.P.E realizou um estudo de *linkage* em todo o genoma, através de *arrays* de 50000 *single-nucleotide polymorphisms* (SNPS), em 2 famílias FCCTX (uma família TSG⁺ e outra família TSG⁻). Estas foram seleccionadas, de entre o grupo de 15 famílias, mencionado anteriormente, por serem as famílias mais informativas uma vez que dispunham de material biológico de vários indivíduos (afectados e não afectados). Deste estudo, foram seleccionadas 3 regiões cromossómicas (13q, 16q, 21q) na família TSG⁺, para as quais se observou evidência de *linkage*. Posteriormente, foi realizada uma análise de *linkage* e LOH recorrendo a marcadores de microssatélite, a qual levou à exclusão da região 16q por não ter revelado *linkage* com a doença (Belo, 2010). A análise de LOH, efectuada nos tumores das 15 famílias FCCTX, permitiu ainda delimitar uma região mínima de LOH de 0,84Mb e de 1,3Mb para as regiões 13q e 21q, respectivamente. No mesmo estudo, observou-se ainda que a LOH no cromossoma 21 se associava aos carcinomas enquanto no cromossoma 13 esta era mais frequente em adenomas. Estes resultados sugerem que na sequência adenoma-carcinoma, os eventos em 13q ocorrem ao nível da iniciação tumoral enquanto os eventos em 21q ocorrem ao nível da progressão tumoral (Pereira, 2013) (Belo, 2010).

A família TSG⁻ foi seleccionada para uma análise de exoma visto não ter apresentado qualquer evidência de *linkage*.

1.5.2.3. Identificação de novas variantes para a susceptibilidade para o FCCTX através de *whole-exome sequencing*

A evolução da sequenciação de nova geração permitiu a exploração de novas ferramentas para o estudo genético de várias doenças. Vários estudos associados ao genoma humano, como a sequenciação de todo o exoma (WES), que envolve a sequenciação de toda a região codificante, têm-se revelado promissores na identificação de mutações patogénicas em genes que não eram associados *à priori* a determinadas doenças. De forma sucinta, inicialmente o DNA é fragmentado e hibridado com sondas, marcadas com biotina, específicas para reconhecer a região de interesse no genoma (todos os exões). Durante os passos de enriquecimentos, estas regiões são capturadas com recuso a esferas magnéticas com estreptavidina, que se ligam às sondas, e a restante parte do genoma é descartada antes da amplificação por PCR da amostra. Por fim, ocorre a sequenciação de nova geração da amostra, seguida por uma análise bioinformática, a qual permite identificar as variantes presentes na região alvo. Actualmente esta técnica tem sido utilizada recorrentemente, tanto na detecção de

mutações somáticas em cancros esporádicos como mutações germinais em cancros hereditários, como o caso do FCCTX (Bamshad *et al.*, 2011) (Rabbani *et al.*, 2014) (Warr *et al.*, 2015).

No seguimento do projecto que visa a identificação de genes candidatos de susceptibilidade para o FCCTX, que está a ser desenvolvido pelo grupo de gastroenterologia da UIPM, foram seleccionadas amostras de DNA genómico de 4 indivíduos da família L56 (TGS⁻) para análise de WES e consequente análise bioinformática, a qual foi efectuada na Erasmus University Medical Center - Center for Biomics (Roterdão, Holanda) numa plataforma *HiSeq* (Illumina). Com esta análise pretendeu-se identificar as variantes partilhadas pelos quatro indivíduos e avaliar a sua patogenicidade e associação com a doença. Uma primeira análise bioinformática revelou um haplótipo partilhado na família envolvendo mutações presentes nos genes *RNF207* (*Ring finger protein 207*), *ZBTB48* (*Zinc finger and BTB domain containing 48*) e *TAS1R1* (*Taste 1 receptor member 1*), localizadas numa região cromossómica que tinha revelado uma pequena sugestão de *linkage* com a doença na família (Saramago, 2014).

De modo a tornar a análise bioinformática mais robusta, foi efectuada a análise de WES (no mesmo local que as anteriores) para um quinto indivíduo desta família. Com esta informação, foi realizada uma nova análise bioinformática, agora para os 5 indivíduos da família L56, na empresa *Bioinf2Bio, Lda* (Porto, Portugal). Os indivíduos seleccionados encontram-se identificados na árvore geneológica (figura 1.7) como individuo 1 (L674), individuo 2 (L801), individuo 3 (L408), individuo 4 (L1794) e individuo 5 (L781).

De modo a maximizar a identificação de mutações germinais potencialmente patogénicas pretendeu-se que a análise bioinformática abrangesse todas as possibilidades e por isso esta foi subdividida em cinco análises, cada uma com uma combinação única dos indivíduos que partilham as variantes (tabela 1.2). Assim, a análise 1 continha as variantes partilhadas por todos os 5 indivíduos, enquanto as restantes análises continham as variantes partilhadas por apenas quatro ou três indivíduos. Nas análises 2 e 3 houve exclusão de um indivíduo (4 e 3, respectivamente), na análise 4 a exclusão dos indivíduos 2 e 4 e na análise 5 a exclusão dos indivíduos 3 e 4. Nestas análises colocou-se a hipótese de os indivíduos 2 e 3 poderem ser potenciais fenocópias por não se poder excluir a hipótese de serem casos esporádicos, embora pouco provável, e por isso não partilharem uma variante hereditária num contexto de uma família FCCTX. O indivíduo 4 também foi considerado como possível caso esporádico visto que, comparativamente com a idade de desenvolvimento de adenomas noutros familiares, aos 49 anos ainda só desenvolveu 2 adenomas.

O resultado das cinco análises bioinformáticas foi fornecido, pela empresa *Bioinf2Bio Lda*, organizado em quatro *outputs* (formato *.xlsx*), os quais contêm informações descritivas sobre as variantes partilhadas em cada análise, segundo os critérios referidos na tabela 1.3. Para o presente estudo, o ponto de partida para a selecção de variantes possivelmente patogénicas que possam justificar o fenótipo FCCTX na família L56 foram os *outputs* 3 e 4. O *output* 4 descreve as variantes contidas no *output* 3 que apresentam alto impacto segundo os *softwares* *SIFT*/*Polyphen*. A selecção

de variantes efectuou-se primeiramente no output 4 por descrever as variantes com maior potencial patogénico. O *output* 3 foi posteriormente incluído para selecção de variantes de modo a evitar a eliminação de variantes que possam ser igualmente importantes, excluídas apenas pelos seus *scores* nestes *softwares*.

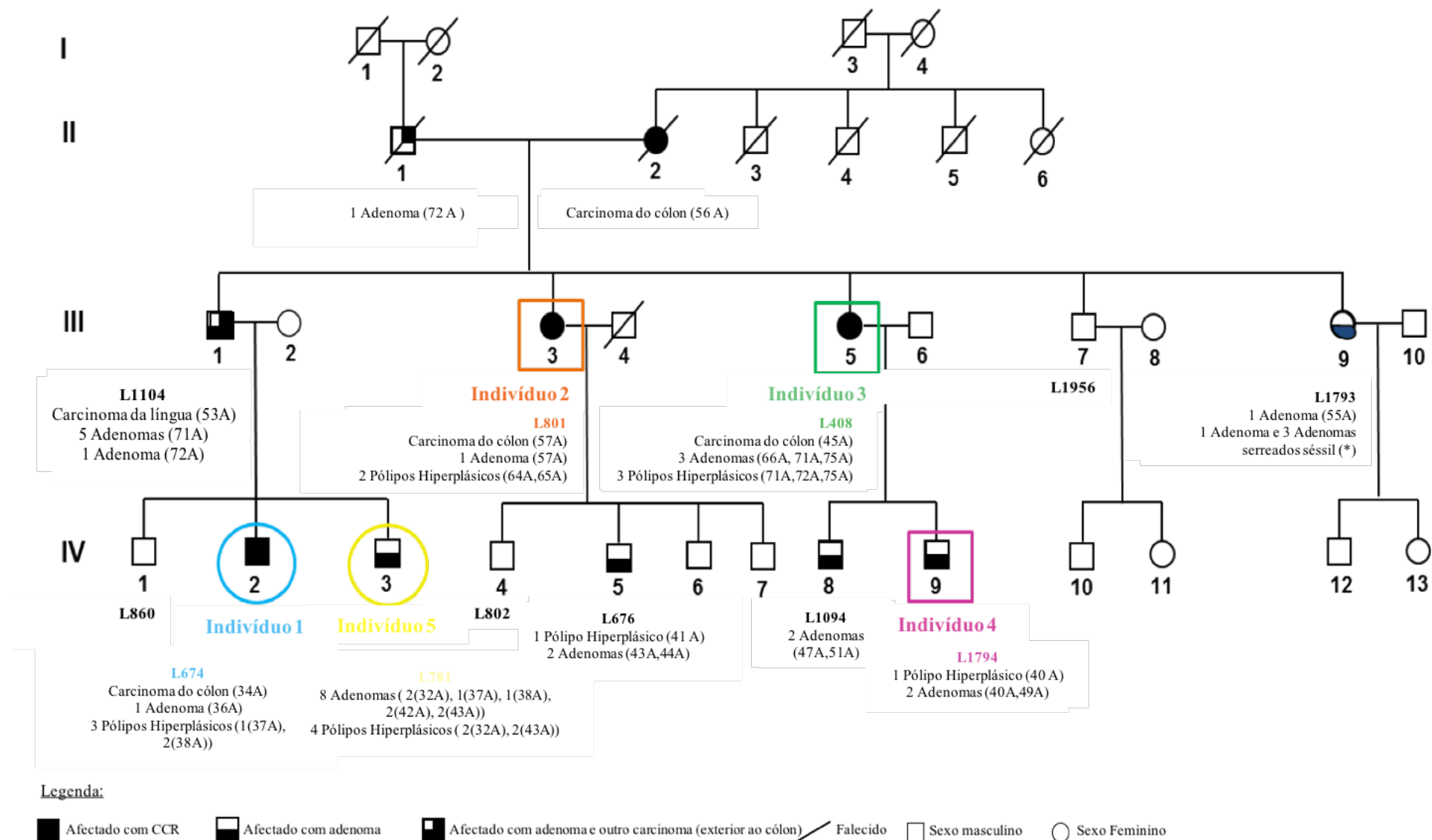


Figura 1.7. Árvore geneológica da família L56. Com cores encontram-se assinalados os indivíduos que foram seleccionados para análise de WES. Apenas para os indivíduos com número de identificação de laboratório existe material biológico disponível. A idade de diagnóstico (em anos, A) e respectiva caracterização fenotípica encontra-se descrita para os indivíduos afetados. Todos os adenomas identificados correspondem a adenomas tubulares com displasia de baixo grau, com a exceção no indivíduo L1104 que apresenta uma adenoma tubular com displasia de alto grau aos 71A. (*) Não foi possível obter-se idade de diagnóstico.

Tabela 1.2. Descrição dos *outputs* obtidos após a análise bioinformática e respetivo número de variantes total obtido.






Análise	Estratégia da análise	Descrição	Nº de variantes
[1]		A análise 1 inclui as variantes partilhadas por todos os indivíduos	19 104
[2]		A análise 2 inclui as variantes partilhadas pelos indivíduos 1,2,3 e 5, excluindo o 4 por ser uma possível fenocópia	3 198
[3]		A análise 3 inclui as variantes partilhadas pelos indivíduos 1,3,4 e 5, excluindo o 2 por ser uma possível fenocópia	2 036
[4]		A análise 4 inclui as variantes partilhadas pelos indivíduos 1,3,e 5, excluindo o 2 e 4 por serem possíveis fenocópias	1 327
[5]		A análise 5 inclui as variantes partilhadas pelos indivíduos 1,2,e 5, excluindo o 3 e 4 por serem possíveis fenocópias	2 254

Tabela 1.3. Descrição da estratégia utilizada nas cinco análises bioinformáticas efectuadas a partir dos resultados obtidos por WES, para os indivíduos seleccionados da família L56 e respetivo número de variantes obtido.

Output	Descrição do output	Nº Variantes Total (Análise 1+2+3+4+5)
[1]	Variantes contidas em regiões codificantes que apresentam um genótipo definido e qualidade de alinhamento e de genótipo assim como número de <i>reads</i> superior a 20	26 301
[2]	Variantes contidas em transcritos codificantes com informação proveniente da base de dados COSMIC	26 301
[3]	Variantes com consequência prevista: <ul style="list-style-type: none"> • <i>frameshift</i>; • variantes em região de <i>splicing</i>; • <i>missense</i>; • variantes aceitador de <i>splicing</i>; • variantes 3'UTR e 5'UTR; • variantes sinónimas; • variantes com ganho de codão STOP ou perda de codão START 	18 642
[4]	Variantes descritas no <i>output</i> 3 com alto impacto segundo os <i>softwares</i> SIFT e Polyphen	604

As análises efectuadas subsequentemente a esta análise bioinformática foram já desenvolvidas no âmbito do presente projecto de mestrado.

2. Objectivos

No global, o presente projecto pretendeu identificar genes candidatos/variantes específicas em genes candidatos que possam estar envolvidos na susceptibilidade para o cancro do cólon e recto familiar do tipo X (FCCTX).

Assim, este projecto teve como primeiro objectivo avaliar o envolvimento do gene *FBXO11* como candidato para a susceptibilidade para o FCCTX através de um estudo mutacional nos seus 22 exões prosseguido por análise de *copy-number* e de expressão por PCR quantitativo.

Seguidamente, com a finalidade de caracterizar dois subgrupos de famílias de FCCTX TGS⁺ e TGS⁻, a nível germinal, foi efectuada uma análise de um painel multigénico incluindo 94 genes associados ao aumento do risco para cancro.

Adicionalmente, pretendeu-se ainda estudar mutações específicas em genes candidatos para a susceptibilidade para o FCCTX, obtidas através de *whole exome sequencing* (WES), avaliando o seu potencial carácter patogénico, de modo a identificar mutações que possam predispor para esta condição hereditária.

3. Material e Métodos

3.1. Material Biológico

A análise de mutações germinais no gene *FBXO11* foi efectuada em 24 amostras de DNA genómico obtido a partir de sangue periférico, 15 de indivíduos índex das famílias FCCTX e 3 de familiares afectados, e ainda 5 de indivíduos HNPCC e/ou que preenchem os CB, cujos tumores apresentavam ausência de expressão do gene *MSH6*, sem mutação germinal identificada nos genes MMR. Para estas mesmas amostras foi também efectuada a análise de *copy-number*, tendo-se utilizado 2 amostras de DNA genómico de indivíduos saudáveis como referência. Em 4 das 5 amostras de indivíduos HNPCC e/ou que preenchem os CB (não foi possível obter RNA de uma das amostras), foi ainda sintetizado DNA complementar (cDNA) a partir de RNA total extraído a partir de sangue periférico, de modo a se proceder à análise de expressão do gene *FBXO11*, tendo-se utilizado 8 amostras de cDNA de indivíduos saudáveis, como controlos normais.

No estudo da caracterização germinal dos subgrupos TGS⁺ e TGS⁻, através da análise de um painel multigénico por NGS, foram incluídas as 15 amostras de DNA genómico obtido a partir de sangue periférico dos indivíduos índex de famílias FCCTX, das quais 10 pertencem ao subgrupo TGS⁺. De acordo com os resultados obtidos para estas famílias, foram seleccionadas 78 amostras, de indivíduos pertencentes a famílias com características semelhantes a estas (43 de indivíduos índex de famílias HNPCC-like e 35 de doentes que preenchiam os CB), de modo a avaliar se os genes identificados nas famílias FCCTX também estariam envolvidos nestas.

Após selecção das variantes, obtidas por WES para a família FCCTX L56, foi efectuada a análise de segregação destas com a doença, em 12 amostras de DNA genómico obtidas a partir de sangue periférico, sendo 9 de indivíduos afectados e 3 de indivíduos não afectados. Para uma das variantes seleccionadas, que se revelou promissora na análise de segregação, foi ainda estudado o seu envolvimento a nível somático, por análise mutacional, de modo a avaliar eventual perda de heterozigotia, e por análise de expressão do gene. Assim, a análise mutacional foi efectuada em 4 amostras de DNA genómico obtido a partir de tecido tumoral (3 de tecido incluído em parafina e 1 de tecido congelado), de 4 indivíduos pertencentes à família L56 e uma amostra de um indivíduo não pertencente à família, como controlo. A análise de expressão do gene foi efectuada em cDNA sintetizado a partir de RNA extraído de uma amostra de tecido tumoral incluído em parafina, de um dos doentes da família L56. Como controlo, foram incluídas amostras de cDNA, sintetizadas a partir de RNA extraído de tecido incluído em parafina, de duas mucosas normais, dois carcinomas e ainda do RNA extraído da linha celular humana de carcinoma do cólon *SW480*.

Todos os indivíduos analisados neste estudo pertencem a famílias seguidas na Clínica de Risco Familiar do Cancro do Cólon e Recto do IPOLFG,EPE. Cada indivíduo assinou um consentimento informado que permitiu efectuar o estudo.

3.2. Métodos

As amostras de DNA e RNA obtidas de sangue periférico, as amostras de DNA obtidas a partir de tecido incluído em parafina e tecido congelado assim como o RNA da linha celular *SW480*, utilizadas neste estudo já tinham sido previamente extraídas.

3.2.1. Extracção de ácidos nucleicos

3.2.1.1. Extracção de DNA genómico e RNA a partir de sangue periférico

A extracção de DNA genómico partir de sangue periférico foi efectuada recorrendo ao *kit* comercial *Gentra Puregene Blood Kit (Qiagen)*, seguida por quantificação do DNA num *NanoDrop 2000 (Thermo Scientific)*, de acordo com as especificações do fabricante. Após a quantificação, o DNA foi diluído em água bidestilada de modo a se obter uma concentração de trabalho de 80 ng/μl e posteriormente armazenado a -20°C.

A extracção de RNA a partir de sangue periférico realizou-se segundo o método descrito para o reagente *TRI Reagent® Solution (Ambion, Life Technologies)*. Após a extracção, as amostras de RNA foram quantificadas num *NanoDrop 2000 (Thermo Scientific)* e sujeitas a uma electroforese em gel de agarose de 2% (p/v) (ponto 3.2.3.1.2.) de modo a avaliar a sua integridade e detectar possíveis contaminações com DNA genómico. Cada amostra foi preparada a partir de 1 μl do RNA extraído, ao qual foram adicionados 9 μl de *Orange G 1x*, tendo sido aplicado o volume total no gel assim como 6,5 μl dos marcadores de peso molecular λ /HindIII (*Fermentas*) e *GeneRuler50bpDNALadder (ThermoFisher Scientific)* (Anexo D). As amostras de RNA foram posteriormente armazenadas a -80°C.

3.2.1.2. Extracção de DNA e RNA a partir de tecido tumoral e mucosa normal incluído em parafina

A extracção de DNA genómico a partir de tecido incluído em parafina foi efectuada recorrendo ao método de fenol-clorofórmio. O DNA extraído foi quantificado em gel de agarose 0,8% (p/v)

(Anexo D), por comparação com amostras de DNA de concentrações conhecidas, tendo sido também avaliada a sua integridade. A preparação da electroforese decorreu como descrito no ponto anterior mas a partir de 1 µl do DNA extraído. Após a quantificação, o DNA foi diluído para uma concentração de trabalho de 80 ng/µl e armazenado a -20°C.

A extracção de RNA, para as amostras L674T, Mucosa normal B e C, foi efectuada a partir de 4 cortes de 10 µm de tecido incluído em parafina utilizando o *kit* comercial *E.Z.N.A.®FFPE RNA Kit Protocol - Xylene Extraction Method (Omega)*. Enquanto que para as amostras Cas638C e A193C foi utilizado o *kit PureLink™ FFPE RNA Isolation Kit (Ambion)*. Após a extracção do RNA, seguiu-se o mesmo procedimento referido em 3.2.1.1..

3.2.1.3. Extracção de DNA a partir de tecido congelado

A extracção da amostra L676P foi efectuada utilizando-se um método de *salting-out* (Miller et al., 1988). A quantificação foi efectuada como descrito no ponto 3.2.1.1.

3.2.1.4. Extracção de RNA a partir de células da linha de carcinoma do cólon SW480

Após obtenção de um *pellet* de células da linha de carcinoma SW480, procedeu-se à extracção do RNA utilizando o *kit TripleXtractor directRNA Kit (GRiSP Research Solutions)*. A quantificação foi efectuada como descrito no ponto 3.2.1.1..

3.2.2. Reacção em cadeia da polimerase (PCR)

A técnica de reacção em cadeia da polimerase (PCR) é um método enzimático *in vitro* que consiste na amplificação de fragmentos específicos de DNA, através da utilização de *primers* que flanqueiam a região de interesse nas duas cadeias nucleótídicas de DNA. A primeira etapa do PCR, denominada por desnaturação, consiste na separação da cadeia dupla de DNA a uma temperatura elevada (94°C). A etapa seguinte consiste no *annealing* dos *primers*, ou seja, na hibridação destes com a sequência complementar na cadeia simples de DNA. A última etapa (extensão) permite a síntese de uma nova cadeia complementar através da acção de uma enzima DNA polimerase. Para além desta enzima, para que a reacção ocorra, é também necessária a presença de desoxinucleótidos (dNTPS) com um grupo 3'-OH livre, o qual permite a ligação de novos nucleótidos e por conseguinte a síntese da nova cadeia de DNA; de uma solução tamponada para manter o pH óptimo à actividade da polimerase e de uma solução com iões magnésio que funciona como cofactor enzimático. A reacção

de PCR é constituída por um número de ciclos que se baseiam na repetição destas três etapas entre 30 a 40 vezes de modo a se obter um elevado número de cópias da cadeia molde.

3.2.2.1. Optimizações das condições de PCR

As condições de PCR foram optimizadas para cada fragmento/exão a amplificar neste estudo, de modo a se obter um fragmento específico, assim como o melhor rendimento possível da reacção. O primeiro passo consistiu no cálculo da temperatura de *annealing* teórica para cada *primer* através da fórmula 1 descrita em baixo:

(Fórmula 1)

$$T_{\text{annealing teórica}} = 2 \times (\text{Número A+T na sequência do primer}) + 4 \times (\text{Número de G+C na sequência do primer})$$

De seguida, procedeu-se a uma titulação da solução contendo iões magnésio, de modo a aferir qual a concentração óptima para se obter um produto de PCR específico, visto que estes funcionam como cofactor enzimático da DNA polimerase. Nos casos que ainda apresentavam inespecificidade, foi necessário diminuir a concentração de magnésio e/ou aumentar a temperatura, de modo a tornar mais específica a hibridação dos *primers* com a sequência alvo e assim, diminuir a amplificação de fragmentos inespecíficos. A concentração de *primer* a utilizar também pode influenciar a especificidade da reacção. Assim, quando se observou a formação de dímeros de *primer*, ou seja, que estes se ligavam preferencialmente entre si, reduziu-se a concentração de *primer* na reacção e/ou aumentou-se a temperatura de *annealing* para tornar a ligação dos *primers* mais específica. Nos casos em que a região alvo apresenta um elevado conteúdo de GC, o que dificulta o processo de desnaturação de cadeia dupla de DNA, foi necessário recorrer a um aditivo, *GC-RICH resolution solution* (5M) pertencente ao *kit GC-RICH PCR System (Roche)*, de modo a melhorar a amplificação do fragmento. Outro tipo de aditivo utilizado foi o DMSO (demitilsulfóxido) com a finalidade de estabilizar o DNA e melhorar a sua desnaturação no primeiro passo da reacção. Para a maioria dos fragmentos as condições de PCR foram optimizadas utilizando a enzima polimerase contida no *kit BiotaqTM (Bioline)*. Porém, para alguns casos particulares houve necessidade de recorrer a outro *kit* de amplificação, *AmpliTa^Q Gold[®] (Thermofisher Scientific)*. A enzima contida neste *kit* tem a particularidade de se encontrar na forma inativa, sendo necessário um “*hot start*” térmico, para iniciar a actividade, o que a torna mais robusta e melhora a especificidade e rendimento da reacção.

3.2.3. Análise do gene *FBXO11* como candidato para a suscetibilidade para o FCCTX

De forma a concluir a análise mutacional do gene *FBXO11*, nas 15 famílias FCCTX, foram analisados os exões 1-6, 9-12, 16, 17 e 20, num total de 10 fragmentos, para os 15 indivíduos índice e 3 familiares. A esta análise, foram adicionadas 6 amostras de DNA genómico de indivíduos cujos tumores apresentavam ausência de expressão do gene *MSH6* e que não tinham mutação identificada nos genes MMR, sendo que para estas, foi efectuada a análise total do gene, ou seja, 22 exões num total de 16 fragmentos.

Em seguida, para as 24 amostras de DNA genómico, foi efectuada análise de *copy number* para o gene *FBXO11* de modo a avaliar possíveis deleções/duplicações através de *Real Time PCR (qPCR)*, tendo sido utilizadas duas amostras de DNA de indivíduos normais como referências. Por fim, para as 6 amostras de indivíduos cujos tumores apresentam ausência de expressão do gene *MSH6*, foi ainda sintetizado cDNA a partir de RNA total, de modo a se proceder à análise de expressão do gene *FBXO11*, por qPCR, tendo sido utilizadas 8 amostras de cDNA de indivíduos saudáveis como controlos normais.

3.2.3.1. Análise mutacional do gene *FBXO11*

3.2.3.1.1. Amplificação do DNA genómico por PCR

Todas as amostras foram amplificadas por PCR para um volume final de 12,5 µl contendo 1 µl de DNA genómico (80 ng) em tubos tipo *ependorf* de 0,2ml. A maioria das reacções de amplificação foram efectuadas utilizando o *kit BiotaqTM (Bioline)*, sendo que para os exões 4, 12, 16 e 17, do gene *FBXO11* foi utilizado o *kit* de amplificação, *AmpliTa^q Gold[®] (ThermoFisher)*. Assim, de um modo geral, para cada reacção foi efectuada uma mistura reacional contendo 1,25 µl de solução tamponada (10x); 1 µl de dNTPs (200 mM, *IllustraTM, GE Healthcare*); quantidade variável de cada *primer* (10 pmol/µl) e de solução de MgCl₂ (50 mM, *Bioline* e 25mM, *ThermoFisher*) de acordo com as condições optimizadas; 0,08 µl de *Taq DNA polymerase* (5 U/µl, *Bioline* e *ThermoFisher*) e água bidestilada para perfazer o volume.

Para a amplificação por PCR dos vários exões do gene *FBXO11* foram utilizados os *primers* descritos por Duan et al., 2012 tendo-se utilizado os *softwares* bioinformáticos *NetPrimer* (<http://www.premierbiosoft.com/netprimer/>) e *UCSC In-Silico PCR* (<https://genome.ucsc.edu/cgi->

bin/hgPcr) para o cálculo da temperatura de *melting* e previsão do tamanho do fragmento esperado, respectivamente (Tabela I do anexo A). As condições de amplificação utilizadas para cada fragmento estão descritas no anexo B (Tabela I), sendo que as reacções de PCR foram executadas num termociclador *Veriti® Thermal Cycler (Applied Biosystems)* de acordo com os programas descritos nas tabelas I e II do anexo C.

3.2.3.1.2. Controlo da eficiência dos produtos amplificados por PCR em gel de agarose 2% (p/v)

A eficiência das reacções PCR foi avaliada através de uma electroforese em gel de agarose, o qual se encontra mergulhado num tampão salino. Assim, esta técnica permite a separação dos fragmentos de DNA ao longo do gel de acordo com os respectivos pesos moleculares, sendo que os fragmentos mais pequenos migram mais que os maiores. Para a visualização destes fragmentos é necessário a utilização de um corante, como o brometo de etídio, que se intercala na cadeia dos ácidos nucleicos e que emite fluorescência após a incidência de radiação ultravioleta.

O controlo da eficiência das optimizações das reacções PCR foi efectuado pela adição de 3 µl de *Orange G 5x* (tampão utilizado para conferir densidade às amostras e assim facilitar a sua deposição nos poços do gel e ainda permitir a visualização da migração durante a electroforese) ao produto amplificado e aplicação da totalidade desse volume nos poços do gel de agarose. No controlo da eficiência das reacções de PCR das amostras de interesse, retirou-se uma alíquota de 3 µl de produto amplificado ao qual se adicionou 7 µl de *Orange G 1x*, tendo-se aplicado os 10 µl resultantes nos poços do gel. Em ambos os casos foi aplicado paralelamente 6,5 µl do marcador de peso molecular *GeneRuler 50bp DNA Ladder (ThermoFisher Scientific)*, de forma a confirmar o tamanho do fragmento amplificado. Deste modo, a electroforese foi realizada numa tina horizontal utilizando um gel de agarose 2% (p/v) em TBE 1x, ao qual foi aplicado uma diferença de potencial de 140 V durante cerca de 25-30 minutos. Após a electroforese o gel foi exposto a radiação ultravioleta num transiluminador *BioDocAnalyze (Biometra)*. Todas as soluções utilizadas neste ponto encontram-se descritas no anexo D.

3.2.3.1.3. Purificação enzimática dos produtos amplificados por PCR

A purificação enzimática do produto amplificado por PCR consistiu na eliminação de resíduos de reagentes, como excesso de dNTPs ou *primers* que não hibridaram, assim como na purificação do DNA amplificado, através da utilização de duas enzimas: *Thermosensitive Alkaline Phosphatase - FastAP* (1U/µl, *Thermo Scientific*) e *Exonuclease I -ExoI* (20U/µl, *Thermo Scientific*). Para tal, foi efectuada uma mistura reaccional contendo 0,5 µl de ExoI e 1 µl de FastAP por cada produto da

reacção, sendo posteriormente adicionados 1,5µl da mistura a cada tubo tipo *eppendorf* de 0,2ml contendo o produto de PCR. Em seguida, as amostras foram incubadas num termociclador durante 15 minutos a 37°C (temperatura óptima para a actividade das enzimas) seguidos de 15 minutos a 85°C para a inactivação das mesmas.

3.2.3.1.4 .Sequenciação automática

A sequenciação automática de Sanger, baseia-se no princípio da replicação do DNA permitindo determinar a sequência nucleotídica presente no fragmento de interesse. Neste método para a síntese de DNA é necessário não só os dNTPs mas também didesoxirribonucleótidos (ddNTPs) que têm a particularidade de não possuírem o grupo hidroxilo no carbono 3' o qual é necessário para a formação de ligações fosfodiéster entre as bases nucleotídicas adjacentes. Assim, a inexistência deste grupo leva a que a DNA polimerase termine a síntese da nova cadeia após a incorporação de um ddNTPs, resultando vários fragmentos que se diferenciam num nucleotido. Para além disto, cada ddNTPs é marcado com um fluorocromo de cor diferente consoante a base nucleotídica (A,G,C,T) que emite fluorescência quando excitado por um laser, permitindo determinar a sequência do fragmento de DNA.

3.2.3.1.4.1. Reacção de Sequenciação

Para cada fragmento amplificado foi efectuada uma reacção de sequenciação, na qual se utilizou apenas um *primer* (*forward* ou *reverse*), sendo que no caso do gene *FBXO11* sempre que se identificou alguma variante esta foi confirmada com nova reacção, tendo-se utilizado o outro *primer*. A reacção de sequenciação foi efectuada para um volume final de 20 µl em água bidestilada, em tubos tipo *eppendorf* de 0,2ml, contendo uma quantidade de produto de PCR purificado variável, de acordo com o tamanho do fragmento amplificado e da intensidade da banda visualizada após o controlo da eficiência da amplificação em gel de agarose como descrito no ponto 3.2.3.1.2.; 2 µl de *primer* (1,6 pmol/µl) específico para a sequência alvo (*forward* ou *reverse*); 2 µl de solução tamponada *Buffer Sequencing 5x*, *BigDye® Terminator v1.1* (*Applied Biosystems*) para manter o pH constante e quantidade variável de *mix BigDye™ Terminator v1.1 Cycle* (*Applied Biosystems*) constituída pela DNA polimerase, Mg²⁺, dNTPs e ddNTPs, a qual depende do tamanho do fragmento amplificado. A reacção de sequenciação foi realizada num termociclador *Veriti® Thermal Cycler* (*Applied Biosystems*) utilizando o programa descrito da tabela III do anexo C. Os *primers* utilizados para a reacção de sequenciação encontram-se no anexo A (Tabelas I e II).

3.2.3.1.4.2. Precipitação e purificação do DNA após reacção de sequenciação

A precipitação e purificação do DNA, foi realizada até um período máximo de 24 horas após a reacção de sequenciação. Este método baseia-se numa reacção de *salting out*, a qual é recomendado pelo protocolo do *kit* de sequenciação *BigDye® Terminator v1.1 Cycle Sequencing (Applied Biosystems)*, tendo como objectivo a obtenção do DNA purificado e sem contaminantes que pudessem interferir na electroforese capilar. Para tal, foi efectuada uma mistura reaccional contendo etanol absoluto, acetato de sódio e EDTA, sendo que os primeiros dois reagentes são essenciais para neutralizar as cargas do ácido nucleico, reduzir a sua solubilidade e promover a precipitação de DNA, enquanto que o EDTA é fundamental para inactivar a enzima usada na reacção de sequenciação, através da sua capacidade de quelatar o Mg^{2+} . Assim, a esta mistura foi adicionado o volume total do produto da reacção de sequenciação, seguido de incubação à temperatura ambiente. De seguida, as amostras foram centrifugadas a 4°C, o sobrenadante foi descartado e os *pellets* foram lavados para remoção de sais em excesso, por centrifugação, com etanol 70% (v/v). O sobrenadante foi novamente descartado e os *pellets* de DNA foram secos a 37°C num banho seco e posteriormente armazenado a 4°C até se efectuar a electroforese capilar. O protocolo detalhado deste procedimento encontra-se descrito no anexo E.

3.2.3.1.4.3. Preparação das amostras para electroforese capilar no sequenciador automático utilizando o software *Sequencing Analysis*

Aos *pellets* secos de DNA, obtidos após a precipitação e purificação das amostras da reacção de sequenciação, foram adicionados 17µl de formamida HI-DI (*Applied Biosystems*) como solvente para a injeção capilar. Em seguida homogenizou-se em *vortex* e transferiu-se o volume total para uma placa de 96 poços (*Platamax, Axygen*) adequada para o sequenciador automático. A placa foi selada e colocada num termociclador *Veriti® Thermal Cycler (Applied Biosystems)* para desnaturação das amostras a 95°C durante 5 minutos, sendo de seguida incubada em gelo durante 1-2 minutos. Por fim, procedeu-se a uma centrifugação durante 2 minutos a 1200 rpm, para deposição das amostras e eliminação de possíveis bolhas de ar. Em seguida, retirou-se o selo e a placa com as amostras foi colocada no sequenciador automático *ABI Prism™ 3130 Genetic Analyser (Applied Biosystems)*, onde a electroforese capilar decorreu a 50°C e a uma voltagem de 15KV.

3.2.3.1.4.4. Análise de resultados

Os resultados da electroforese capilar foram obtidos sob a forma de electroforectogramas gerados pelo *software Sequencing Analysis 3.4.1 (Applied Biosystems)*. Para identificar as alterações na sequência alvo, a sequência obtida foi comparada com a sequência de DNA de referência obtida da base de dados NCBI, sendo que, para além de toda a sequência codificante teve-se também atenção a possíveis alterações nas 20 bases franqueadoras, a montante e jusante, de cada exão.

3.2.3.2. Real-Time PCR

A técnica de *Real-Time* PCR ou PCR quantitativo (qPCR) é uma técnica sensível, específica e que permite a quantificação de ácidos nucleicos. Esta técnica tem vindo a ser usada nas mais diversas aplicações, desde estudos de expressão de mRNA, estudos de *copy-number* (quantificação de DNA genómico) ou de variantes específicas. O seu procedimento tem como base o princípio do PCR convencional, mas com a particularidade de usar moléculas fluorescentes na reacção de amplificação capazes de se intercalarem ou se ligarem à cadeia do DNA, o que torna possível a obtenção de uma razão directamente proporcional entre o sinal de fluorescência e a quantidade de produto presente na reacção. As moléculas fluorescentes dividem-se em dois grupos: as moléculas que intercalam com a cadeia dupla do DNA (exemplo, *SYBR® Green*) ou moléculas marcadas com fluorocromos que se ligam a uma região específica na cadeia simples de DNA através da complementariedade de bases (exemplo, *TaqMan® probe*). O *software* informático do instrumento onde decorre a amplificação por PCR quantitativo, constrói um gráfico de amplificação, semelhante ao ilustrado na figura 3.1, tendo por base a emissão de fluorescência detectada ao longo do número de ciclos. No gráfico de amplificação é possível observar as três fases que constituem o qPCR: a fase limiar (*threshold*), a fase exponencial e a fase estacionária ou *plateau*.

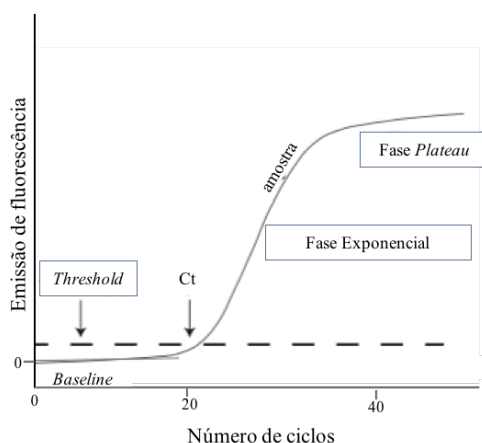


Figura 3.1. Gráfico de amplificação para uma amostra por qPCR e representação das suas respectivas fases. (adaptado de:Arya et al.,2005)

Para cada amostra é determinado, pelo *software* do equipamento, o ponto a partir do qual a fluorescência gerada ultrapassa o *threshold*, em função da *baseline* (limiar mínimo de detecção de fluorescência do instrumento), sendo esse ponto denominado por *threshold cycle* (Ct).

A técnica de qPCR, permite a realização de dois métodos de quantificação: quantificação absoluta ou quantificação relativa. A quantificação das amostras de forma absoluta consiste na determinação da concentração de amostras desconhecidas tendo em conta uma recta de calibração. Inicialmente são efectuadas diluições seriadas de uma amostra conhecida e determinado o Ct para cada diluição. A recta de calibração é desenhada através de uma relação linear dos valores de Ct determinados e o logaritmo da concentração das diluições seriadas. Na quantificação relativa, para além do gene alvo, é necessário um gene *housekeeping* (expresso constitutivamente e de igual forma em todas as células) e uma amostra de referência. Durante o qPCR ocorre a amplificação para cada amostra (alvo e de referência) do gene alvo e do gene *housekeeping*. Assim, neste método, a quantificação consiste na comparação dos valores de Ct entre as amostras do gene alvo contra os Ct das amostras para o gene controlo.

3.2.3.2.1. Optimização das condições da reacção de qPCR e cálculo da sua eficiência

A eficiência da reacção do qPCR depende dos valores de Ct obtidos de acordo com a concentração da amostra e do gene em estudo. Para o cálculo da eficiência, efectuaram-se 5 diluições seriadas em água bidestilada, de uma amostra de referência de DNA genómico ou cDNA, com concentração conhecida. Essas diluições foram efetuadas para as concentrações de 40, 20, 10, 5 e 2,5 ng/μl e 60, 30, 15, 7,5 e 3,25 ng/μl, caso se tratasse da optimização da reacção para a análise de *copy-number* ou de expressão, respectivamente. De seguida, procedeu-se à amplificação por qPCR, em placas de 96 poços (ABI PRISM™ Optical 96-Well Reaction Plate), utilizando o *kit Power SYBR® Green PCR Master Mix* (5mL, *Applied Biosystems*) ou *KAPA SYBR® FAST qPCR Kit Master Mix Universal* (*KapaBiosystems*), utilizando *primers* específicos, para cada gene alvo e *housekeeping*, com as diferentes diluições. Cada diluição foi efectuada em duplicado e calculada a média de Ct, sendo esta utilizada para construir uma recta através de uma relação linear com o logaritmo da concentração da amostra. O declive da recta permitiu então calcular a eficiência da reacção através da seguinte equação:

(Equação 1)

$$\text{Eficiência (E)} = 10(-1/\text{declive})$$

A eficiência calculada permitiu quantificar as amostras para os estudos de análise de expressão e de *copy-number* através da equação 2:

(Equação 2)

$$ratio = \frac{Eficiência\ do\ gene\ alvo^{\Delta Ct_{amostra}}}{Eficiência\ do\ gene\ housekeeping^{\Delta Ct_{Referência}}}$$

Em que as eficiências são calculadas pela equação anterior e os ΔCt são calculados através da equação 3:

(Equação 3)

$$\Delta Ct = Ct_{Gene\ housekeeping} - Ct_{Gene\ Alvo}$$

Todas as reacções foram realizadas para um volume final de 15µl. Com o kit *Power SYBR® Green PCR Master Mix* (5mL, *Applied Biosystems*) foi efectuada uma mistura reacional diferente para cada gene em estudo, constituída por 0,75µl de cada *primer* (*foward e reverse*) à concentração de optimização, 7,5µl de *Power SYBR® Green PCR Master Mix* (2x) e 4µl de água bidestilada. Após a distribuição da mistura reacional em cada poço da placa, adicionou-se 2µl de DNA genómico ou cDNA, para optimização de análise de *copy-number* ou expressão, respectivamente, excepto no controlo negativo onde se adicionou igual volume de água bidestilada. Com o kit *KAPA SYBR® FAST qPCR Kit Master Mix Universal* (*KapaBiosystems*), a reacção foi similar, ou seja, foi efectuada uma mistura reacional constituída por 0,75µl de cada *primer* (*foward e reverse*) à concentração de optimização, 7,5µl *qPCR Master Mix* (2x), mas adicionou-se apenas 3,6 µl de água bidestilada devido à adição de 0,4 µl de *ROX Reference Dye High*, necessária para tornar a reacção com o *SYBR® Green I dye* presente na *master mix* deste kit, compatível com o aparelho usado. As reacções de qPCR foram realizadas no *7900HT Sequence Detection System* (*Applied Biosystems*) com o programa descrito no Anexo C (Tabelas V e VI) sendo a análise dos resultados efetuada pelo *software* SDS 2.4 (*Applied Biosystems*).

3.2.3.2.2. Análise de *copy number* para o gene *FBXO11*

A análise de *copy-number* para o gene *FBXO11* foi realizada para as mesmas 24 amostras de DNA genómico em que foi efectuada a análise mutacional e 2 amostras de DNA de indivíduos normais, como referência.

Para esta análise foram utilizados dois pares de *primers* específicos, com localizações diferentes no gene *FBXO11* (*FBXO11_probe 1* e *FBXO11_probe 2*), e como gene *housekeeping* uma região no cromossoma 12 (6585217pb a 6585317pb), retirados do estudo de *copy-number* realizado por Duan e

os seus colaboradores (Duan et al., 2012). As sequências dos *primers* utilizados, bem como a sua localização encontram-se descritas na tabela III do anexo A. Após a optimização, as reacções de qPCR foram efectuadas em triplicado, como descrito no ponto anterior (3.2.3.2.1.), utilizando o *kit Power SYBR® Green PCR Master Mix* (5mL, *Applied Biosystems*) e 2 µl de DNA (10 ng/µl) de cada amostra por par de *primers* em análise. As condições de amplificação encontram-se descritas na tabela II do anexo B.

A reacção foi efectuada no *7900HT Sequence Detection System* (*Applied Biosystems*) com o programa descrito no Anexo C (Tabela V) sendo a análise dos resultados efectuada pelo *software* SDS 2.4 (*Applied Biosystems*). Após a reacção, os valores de Ct adquiridos pelo *software* foram importados para uma folha de cálculo de Excel e a análise de *copy-number* foi realizada de acordo com as equações 1, 2 e 3. A média dos Ct utilizada foi realizada a partir dos triplicados de cada reacção. A análise estatística foi realizada no *software GraphPad Prism version 5.03*, utilizando a média e o desvio padrão dos triplicados de cada experiência.

3.2.3.2.3. Análise de expressão do gene *FBXO11*

Para os 5 indivíduos incluídos na análise do gene *FBXO11*, cujos tumores apresentam ausência de expressão do gene *MSH6*, foi colocada a hipótese desta ausência ser devida ao mecanismo de expressão *antisense* presente nos genes *FBXO11* e *MSH6* (ponto 1.4). Para tal, foi sintetizado cDNA de 4 destas amostras (para uma das amostras não foi possível realizar a análise por falta de disponibilidade de material biológico) e 8 controlos normais (ponto 3.2.3.2.3.1.), de modo a se proceder à análise de expressão do gene *FBXO11*. Todas as reacções de qPCR foram realizadas conforme descrito no ponto anterior, sendo que neste caso, em cada reacção foram utilizados 2µl de cDNA (15 ng/µl) de cada amostra, por gene analisado. Como controlo foi utilizado o gene *SDHA* (*sucinato desidrogenase subunidade A*). As sequências dos *primers* utilizados, bem como a sua localização encontram-se descritas na tabela III do anexo A. As condições de amplificação encontram-se descritas na tabela II do anexo B. A reacção foi efectuada no *7900HT Sequence Detection System* (*Applied Biosystems*) com o programa descrito no Anexo C (Tabela VI) sendo a análise dos resultados efectuada pelo *software* SDS 2.4 (*Applied Biosystems*).

3.2.3.2.3.1. Reacção de Transcrição Reversa (RT-PCR) para síntese de cDNA

A reacção de RT-PCR tem como finalidade a síntese de cDNA a partir de um molde de RNA. Este procedimento está dividido em duas fases: a primeira em que são utilizados *primers random* que se ligam às cadeias de mRNA para que a enzima transcriptase reversa consiga converter o RNA em cDNA e a segunda, em que ocorre a amplificação da dupla cadeia de cDNA, por PCR. A reacção foi realizada para um volume final de 20 µl, em tubos tipo *ependorf* de 0,2ml, sendo composta por duas

misturas reaccionais. A primeira foi realizada para um volume final de 7,75 µl contendo 0,5 µl de hexâmeros (3µg/µl, *Roche*) que funcionam como *primers random*, volume equivalente a 1 µg de RNA (dependendo da quantificação da amostra) e restante volume em água bidestilada tratada com dietilpírocarbonato (DEPC, *Merck*), para inactivação de RNases. Estes foram incubados a 70°C durante 10 minutos num termociclador e colocados, de seguida, em gelo de modo a se adicionar a segunda mistura reacional, constituída por: 4µl de tampão *5x First Strand Buffer* (*Invitrogen*) para manter o pH da reacção constante; 4µl de dNTPs (100mM cada, *Illustra™ GE Healthcare*); 2µl de ditiotreitól (DTT) (0,1M, *Invitrogen*) como estabilizador enzimático; 1µl de *Superscript® II Reverse Transcriptase* (200U/µl, *Invitrogen*), enzima que sintetiza o cDNA a partir de um RNA molde; 0,75µl de *RnaseOut™ Recombinant Ribonuclease Inhibitor* (40U/µl, *Invitrogen*), para a inactivação de possíveis RNases presentes na reacção e 0,5 µl de água bidestilada tratada com DEPC de modo a perfazer um volume de 12,25 µl. Em seguida, os tubos foram novamente colocados no termociclador para finalizar a reacção de acordo com o programa descrito na Tabela IV do anexo C. As amostras de cDNA sintetizadas foram posteriormente guardadas a -20°C.

3.2.4. Caracterização a nível germinal, das famílias FCCTX (subgrupos TSG⁺ e TSG⁻) através da análise de um painel multigénico por *Next Generation Sequencing*

As amostras de DNA genómico dos 15 indivíduos índice das famílias FCCTX, foram analisadas recorrendo a um painel comercial contendo 94 genes associados à susceptibilidade para o cancro, nomeadamente, ao cancro do cólon e recto, por NGS. Uma vez que estas famílias não apresentam mutações nos genes MMR e *MUTYH*, fomos pesquisar a presença de mutações nos restantes genes do painel de modo a podermos caracterizar os subgrupos TSG⁺ e TSG⁻. Paralelamente, para as mesmas amostras foi efectuada a análise de grandes deleções/duplicações por *MLPA* (*multiplex ligation-dependent probe amplification*) nos genes associados a síndromes hereditárias de cancro colorectal, de acordo com as instruções do fabricante (*MRC-Holland*).

Posteriormente, foram adicionados os resultados obtidos pela análise do mesmo painel multigénico, relativamente aos genes identificados nas famílias FCCTX, para um conjunto de 78 amostras de DNA genómico (43 de indivíduos índice de famílias HNPCC-like e 35 de doentes que preenchem os CB) sem mutação identificada nos genes MMR e *MUTYH*.

A análise mutacional por NGS e de *MLPA* (procedimento experimental e análise de resultados) foi realizada pelo grupo de gastroenterologia da UIPM.

3.2.4.1. Análise mutacional utilizando um painel multigénico e selecção das variantes possivelmente patogénicas

A análise mutacional foi efectuada a partir de cerca de 50 ng de DNA genómico de cada uma das amostras, utilizando o *kit* comercial *Trusight Cancer Panel* (Illumina) de acordo com as especificações do fabricante e realizada na plataforma *MiSeq* (Illumina). A anotação das variantes detectadas, foi efectuada recorrendo ao *software Isaac* (v2.1) (Illumina) e os resultados visualizados com o *software VariantStudio* (v2.3) (Illumina). A anotação das variantes é efectuada em relação ao genoma humano GRCh37 (NCBI) e inclui a análise *in silico* com os *softwares SIFT* e *Polyphen-2*, para previsão do impacto que a alteração de aminoácidos possa ter ao nível proteico; da frequência descrita na população nas bases de dados *1000 genomes* e *Exome Variant Server* (EVS), assim como se as variantes se encontram descritas nas bases de dados *Cosmic* (*Catalogue of Somatic Mutations in Cancer*) e *CliVar* (NCBI).

Para a selecção das variantes possivelmente patogénicas, foi realizada a confirmação da análise *in silico* obtida com os *softwares SIFT* e *Polyphen-2*, utilizando a ferramenta *Variant Effect Predictor* - VEP (http://grch37.ensembl.org/Homo_sapiens/Tools/VEP), à qual foi acrescentada a análise com o *software Mutation taster* (<http://www.mutationtaster.org/>) que permite avaliar o impacto provocado por qualquer tipo de mutação pontual no genoma. Para além da confirmação dos resultados obtidos, nas bases de dados referidas anteriormente, fomos ainda complementar a anotação com a frequência descrita na população, para cada variante, na base de dados *Exome Aggregation Consortium* - ExAC (versão 0.3, <http://exac.broadinstitute.org/>), se estas se encontravam descritas nas bases de dados *Leiden Open Variantion Database* – LOVD (<http://www.lovd.nl/3.0/home>) e *The Human Gene Mutation Database* - HGMD (<http://www.hgmd.cf.ac.uk/ac/index.php> - public database) e com a informação obtida através de pesquisa em artigos científicos da especialidade (Pubmed, NCBI). Deste modo, foram seleccionadas como possivelmente patogénicas, todas as variantes cuja frequência descrita na população fosse inferior a 0,1% (ExAC) e simultaneamente pela análise *in silico* obtivessem um dos seguintes resultados (tabela 3.1):

Tabela 3.1. Critérios usados para a selecção das variantes como possivelmente patogénicas, utilizando os *softwares* VEP (*SIFT* e *Polyphen-2*) e *Mutation Taster*.

Critérios para selecção de variantes possivelmente patogénicas	<i>SIFT</i>	<i>Polyphen-2</i>	<i>Mutation Taster</i>
	deleterious	probably damaging	disease causing
	deleterious	possibly damaging	disease causing (score \geq 70)
	tolerated	probably damaging	

Para o caso das amostras dos indivíduos pertencentes às famílias *HNPCC-like* ou que preenchiam os CB, os critérios de selecção das variantes foram os mesmos, mas cingiram-se apenas a mutações

em genes específicos associados à reparação do DNA, nomeadamente a recombinação homóloga (HR) e a reparação por excisão de nucleótidos (NER).

3.2.4.1.1. Desenho de *primers* específicos para a análise das várias variantes presentes em estudo

Para cada variante foram desenhados *primers* específicos de modo a amplificar a região de interesse. Deste modo, seleccionou-se uma sequência com cerca de 18-22 nucleótidos para cada região/exão de interesse, a partir da sequência de referência de cada gene (GRCh37, NCBI), de modo a que esta flanqueasse a variante seleccionada. Esta sequência foi inserida no *software NetPrimer* (<http://www.premierbiosoft.com/netprimer/>), o qual permite obter informações relativas a determinados parâmetros da sequência seleccionada, nomeadamente a temperatura média de desnaturação (T_m), a percentagem de GC na sequência, assim como avaliar a possibilidade das sequências formarem estruturas secundárias geradas por interacções moleculares entre si (*self-dimer*) ou entre os dois *primers* (*cross-dimer*). Assim, foram seleccionadas as sequências de *primers* em que a distribuição de nucleótidos fosse de tal forma que a percentagem de GC correspondesse a cerca de 40-55% da mesma, uma vez que este valor influencia a T_m . Sendo importante que tanto a percentagem de GC como a T_m fossem semelhantes entre o par de *primers* (*primer forward* e *primer reverse*) escolhido. Foi tida também em consideração a especificidade de *annealing* dos *primers* com a sequência alvo, tendo-se optado, sempre que possível, pelos *primers* que não formassem *self-dimers* ou *cross-dimers*, uma vez que estas podem contribuir para a diminuição ou ausência de amplificação. Após o desenho e selecção do melhor par de *primers* para a amplificação de cada variante, recorreu-se ainda a mais dois *softwares*: *PCR in silico* (UCSC Genome Bioinformatics: <https://genome.ucsc.edu/cgi-bin/hgPcr>) de modo a confirmar o tamanho do fragmento obtido, o qual se pretendeu que não fosse superior a 600 bp, e *Primer-Blast* (NCBI: <https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) para avaliar a especificidade da sequência do *primer* no genoma.

3.2.4.1.2. Confirmação das variantes seleccionadas, por sequenciação de Sanger

Para a confirmação de cada variante seleccionada, foram optimizadas as condições de PCR para cada par de *primers* desenhado, seguido da amplificação da amostra em questão e posterior sequenciação por Sanger como descrito anteriormente (pontos 3.2.2.1, 3.2.3.1.1 – 3.2.3.1.4.4), sendo que neste caso a análise de resultados se cingiu apenas à localização da variante em questão. Esta confirmação foi apenas efectuada para as variantes identificadas nos 15 indivíduos índice das famílias FCCTX. As sequências dos *primers* desenhados para a confirmação das variantes resultantes da análise do painel multigénico, assim como as suas características e as amostras em que foram identificadas estão descritas na tabela IV do anexo A. As condições de amplificação utilizadas para cada fragmento estão descritas no anexo B (tabela III), sendo que as reacções de PCR foram

executadas num termociclador *Veriti® Thermal Cycler (Applied Biosystems)* de acordo com os programas descritos nas tabelas I e II do anexo C.

3.2.5. Identificação de genes candidatos para a susceptibilidade para o cancro do cólon e recto familiar do tipo X (FCCTX)

Como referido anteriormente, de forma a identificar genes candidatos para a susceptibilidade para o FCCTX, foi efectuada a sequenciação do exoma em 5 familiares afectados da família L56 e subsequente análise bioinformática para anotação das variantes identificadas e estratificação de acordo com a possível patogenicidade. De seguida, pretendeu-se refinar esta análise, de modo a se seleccionar as variantes candidatas para o estudo de segregação com a doença.

3.2.5.1. Selecção de variantes identificadas através da análise de *whole exome sequencing* na família L56

De modo a seleccionar as variantes potencialmente patogénicas, partilhadas pelos indivíduos afectados da família L56, para posterior estudo de segregação, foram estabelecidos critérios de exclusão (figura 3.2), os quais foram aplicados às variantes descritas nos *outputs* 3 e 4.

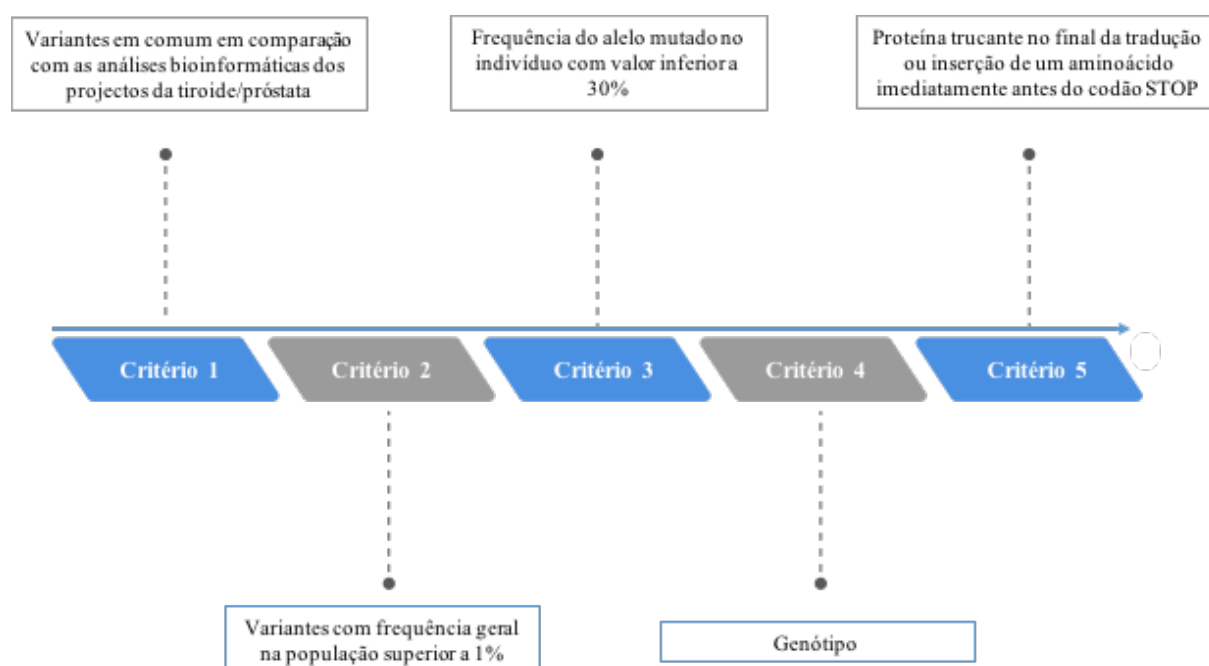


Figura 3.2. Critérios de exclusão utilizados para a selecção das variantes para estudo de segregação com a doença na família L56.

O primeiro critério consistiu na exclusão de variantes comuns, obtidas pela análise bioinformática, às três famílias de cancro familiar (cólon, tireoide e próstata) analisadas por WES de acordo com o projecto referido anteriormente (ponto 1.6). Com o segundo critério pretendeu-se eliminar as variantes que apresentavam frequências gerais na população com valor superior a 1% (por se tratarem de possíveis polimorfismos), seguido pela exclusão de variantes com frequência do alelo mutado, em dois ou mais indivíduos, com valor inferior a 30%, os quais podem representar possíveis artefactos que ocorreram durante a sequenciação. O critério seguinte, teve em conta os genótipos das variantes identificadas nos indivíduos da família, ou seja, se estas eram detectadas em homozigotia ou heterozigotia. Assim, e tendo em conta a transmissão dominante de CCR na família, as variantes identificadas como homozigóticas em todos os indivíduos ou na sua maioria, foram excluídas. Por fim, o último critério serviu para descartar as variantes truncantes cujo codão de terminação se localizasse no final da proteína, por darem origem a proteínas truncadas com tamanho semelhante à *wild-type*, ou seja, com um baixo potencial patogénico. Após a aplicação destes critérios aos dois *outputs* (3 e 4), obteve-se um total de 296 variantes, às quais foram adicionadas quatro variantes resultantes de uma análise bioinformática anterior a este estudo.

De seguida, procedeu-se a uma análise *in silico* para a totalidade das variantes seleccionadas, de modo a tentar restringir este número e ajudar à previsão da patogenicidade das mesmas. As frequências gerais na população foram confirmadas através das bases de dados *Ensembl* (<http://www.ensembl.org/index.html>), ExAC e EVS (versão 0.0.15, <http://evs.gs.washington.edu/EVS/>). Os valores de *SIFT/Polyphen-2*, foram validados através do *software* VEP. Para as variantes sinónimas foi ainda realizada uma análise adicional, com o *software* *Mutation Taster*, de modo a prever possíveis consequências a nível de *splicing*. Por último, foi efectuado um levantamento sobre as funções das proteínas, codificadas pelos genes contendo as variantes seleccionadas, no envolvimento tumoral.

Assim, após esta análise, foram seleccionadas 41 variantes todas em genes diferentes, para análise de segregação com a doença na família.

3.2.5.1.1. Análise de segregação das variantes seleccionadas com a doença, na família L56

Esta análise foi iniciada com o desenho de *primers* específicos para cada uma das variantes, como descrito em 3.2.4.1.1. De seguida, procedeu-se à amplificação por PCR, seguido de sequenciação de Sanger, de cada um dos 47 fragmentos/exões, para as 12 amostras de DNA genómico da famílias L56 (9 de indivíduos afectados e 3 de indivíduos não afectados), como descrito anteriormente (pontos 3.2.2.1, 3.2.3.1.1 – 3.2.3.1.4.4), sendo que neste caso a análise de resultados se cingiu apenas à localização da variante em questão. A maioria das reacções de amplificação foram efectuadas utilizando o *kit* *BiotaqTM* (*Bioline*). Porém, para as variantes identificadas nos genes *TM4SF5* e *MAP6*, foi utilizado o *kit* de amplificação, *AmpliTaQ Gold[®]* (*Thermofisher*). A sequência dos *primers*

desenhados para cada uma das variantes seleccionadas, assim como as suas características encontram-se descritas na tabela V do anexo A. As condições de amplificação utilizadas para cada fragmento estão descritas no anexo B (tabela IV), sendo que as reacções de PCR foram executadas num termociclador *Veriti® Thermal Cycler (Applied Biosystems)* de acordo com os programas descritos nas tabelas I e II do anexo C. Para os genes *CACNA1S* e *GIMAP1* ainda não se obteve as condições óptimas de PCR.

3.2.5.1.1.1. Análise mutacional somática da variante p.Arg645Trp do gene *MTMR3* na família L56

Um estudo realizado por Zheng et al., 2014 em linhas celulares humanas de cancro colorectal, observou que o gene *MTMR3* (*Myotubularin related protein 3*) influenciava os processos de apoptose e proliferação das células de cancro colorectal. Assim, a variante p.Arg645Trp do gene *MTMR3* revelou-se promissora na susceptibilidade para FCCTX, tendo sido seleccionada para uma análise mutacional somática de modo a avaliar a existência de uma eventual perda de heterozigotia. Caso existisse, a variante seria detectada em homozigotia nas amostras de tumor dos indivíduos portadores da variante a nível germinal. Esta análise foi realizada em 4 amostras de DNA genómico, de indivíduos da família L56, para os quais existia material biológico disponível (tabela 3.2). Foi ainda adicionada uma amostra de DNA de um carcinoma de um indivíduo não pertencente a esta família e portanto, não portador da variante germinal.

Tabela 3.2. Caracterização das amostras de DNA dos indivíduos da família L56 incluídas na análise mutacional somática do gene *MTMR3*.

Amostra de DNA genómico	Amostra de DNA de tecido	Caracterização histológica do tecido
L674	Cas 995	Carcinoma
L676	L676P	Pólipo hiperplásico
L781	Cas 987	Adenoma tubular
L801	Cas 3575	Carcinoma

Para esta análise não foi possível utilizar o par de *primers* desenhado anteriormente para o gene *MTMR3* (tamanho do fragmento amplificado - 406bp), uma vez que o tratamento a que o tecido é sujeito até à inclusão em parafina, provoca a fragmentação do DNA, o que inviabiliza a amplificação de fragmentos com tamanho superior a 180-200bp. Para tal, foi necessário o desenho de *primers* específicos (como descrito em 3.2.4.1.1), mas que amplificassem um fragmento de menor tamanho (Anexo A, tabela VI).

A análise somática foi efectuada como descrito anteriormente (pontos 3.2.2.1, 3.2.3.1.1 – 3.2.3.1.4.4) utilizando o kit de amplificação de PCR AmpliTaq Gold® (Thermofisher). As condições de amplificação encontram-se descritas na tabela V do anexo B e as reacções foram realizadas num

termociclador Veriti® Thermal Cycler (Applied Biosystems) de acordo com o programa descrito no anexo C (Tabela II).

3.2.5.1.1.2. Análise de expressão do gene *MTMR3* em tumores por *Real-Time* PCR

Num estudo realizado por (Oppelt et al., 2014) a proteína MTMR3 e o seu alvo PtdInsP5 foram descritos como elementos participantes na migração celular em CCR através da interacção com a proteína RAC1 (*Rho family, small GTP binding protein rac1*). Visto que a isoforma alternativa desta proteína (RAC1b) também já foi descrita como interveniente na tumorigénese nas células do epitélio cólico, estes três genes foram seleccionados para uma análise de expressão em carcinomas do cólon desta família, de modo a avaliar um possível interacção da variante no gene *MTMR3* com a expressão do mesmo ou eventualmente do *RAC1* ou *RAC1b* (Matos e Jordan, 2008).

Para os genes *MTMR3*, *RAC1* e *RAC1b* foram desenhados *primers* específicos de acordo com o descrito anteriormente (ponto 3.2.4.1.1), com a particularidade destes não poderem amplificar um fragmento superior a 200bp, terem uma percentagem em GC de 50% e o valor de T_m ser aproximadamente de 60°C. Além disso, de modo a tornar os *primers* específicos para a análise de cDNA, pelo menos um dos *primers* de cada par foi desenhado de modo a hibridar com uma sequência de junção exão-exão do gene em estudo.

A análise de expressão do gene *MTMR3*, foi efectuada para um dos indivíduos afectados da família L56, numa amostra tumoral (L674T) obtida a partir de tecido incluído em parafina. Para tal, foi sintetizado cDNA desta amostra, assim como de duas amostras de mucosa normal e de dois carcinomas, que foram utilizadas como controlos, como descrito no ponto 3.2.3.2.3.1.. As reacções de qPCR para os genes *MTMR3*, *GADPH* e *RAC1b* foram efectuadas utilizando o *kit KAPA SYBR® FAST qPCR Kit Master Mix Universal* (KapaBiosystems), enquanto que para o gene *RAC1* foi utilizado o *kit Power SYBR® Green PCR Master Mix* (5mL, Applied Biosystems). Todas as reacções foram efectuadas em triplicados e de acordo com o modo descrito em 3.2.3.2.1. Os *primers* específicos usados para cada reacção estão pormenorizados no anexo A (tabela VII) e as condições de qPCR encontram-se descritas no anexo B (tabela VI). A reacção foi efectuada no *7900HT Sequence Detection System* (Applied Biosystems) com o programa descrito no Anexo C (Tabela VI) sendo a análise dos resultados efectuada pelo *software* SDS 2.4 (Applied Biosystems).

4. Resultados

4.1. Envolvimento do gene *FBXO11* na susceptibilidade para FCCTX

4.1.1. Análise Mutacional

A análise mutacional, efectuada para os 22 exões codificantes do gene *FBXO11*, permitiu identificar 7 variantes em heterozigotia em 8 indivíduos da série analisada que incluiu indivíduos index de famílias FCCTX distintas e indivíduos com CCR cujo tumor apresentava ausência de expressão da proteína MSH6. (figura 4.1)

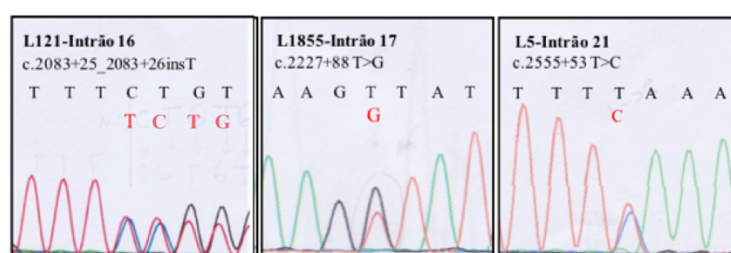


Figura 4.1. Excertos de electroforetogramas representativos de 3 das 7 variantes identificadas. No canto superior esquerdo está indicada a amostra, o exão onde a variante foi detectada e a sua caracterização; as sequências de DNA a preto (sentido *forward*) correspondem à sequência normal naquela região enquanto que a vermelho se encontram as bases/sequência alteradas.

As variantes foram observadas em regiões distintas do gene, desde regiões exónicas (3), intrónicas (2) e região 3'UTR (2), encontrando-se todas descritas na base de dados NCBI. Assim, identificaram-se as seguintes variantes (Tabela 4.1): c.376 A>T (exão 2), c.1399-107_1399-106insAA (intrão10), c.1578A>G (exão 11), c.1398+13_1398+16delATTT (intrão 10), c.2083+25_2083+26insT (intrão 16), c.2227+88T>G (intrão 17-3'UTR) e c.2555+53T>C (intrão 21-3'UTR). Com a excepção da alteração c.2227+88T>G que foi identificada em dois indivíduos, as restantes variantes foram identificadas em apenas um. Na análise mutacional iniciada por Duarte,2015, foram detectadas duas variantes, uma no intrão 17 (c.789+37C>T) e a outra no intrão 12 (c.1451-116 >T). Toda a análise mutacional está resumida na tabela I, do anexo G. As variantes identificadas estão resumidas na tabela 4.1.

No caso de 4 das 7 variantes, pela sua frequência deverão tratar-se de polimorfismos. A variante *missense* c.376 A>T, apesar da frequência baixa na população europeia, apresenta uma frequência alélica elevada na população mundial (2,4%) e está classificada nos *softwares SIFT/Polyphen* como *tolerated/benign*, não devendo assim ser patogénica. A variante sinónima c.1578A>G é classificada como polimorfismo pelo software *Mutation Taster*, e para além de não implicar uma alteração de aminoácido também não implica alteração a nível de *splicing*. A variante descrita para o intrão 10

(c.1398+13_1398+16delATTT) também está classificada como polimorfismo pelo *software Mutation Taster* e revelou elevada frequência alélica tanto mundial (16%) como europeia (27%). Apenas a variante c.2555+53 T>C revelou frequência inferior a 1% (0,7%), mas a sua localização não sugere patogenicidade.

Desta forma, não foi identificada qualquer mutação potencialmente patogénica quer nos indivíduos das 15 famílias FCCTX, quer nos indivíduos cujo tumor apresentava ausência de expressão da proteína MSH6 e que por isso podia estar associada a alguma alteração no gene *FBXO11*, como proposto anteriormente.

Tabela 4.1. Caracterização das variantes obtidas na análise mutacional dos 22 exões codificantes do gene *FBXO11*, para as 15 famílias FCCTX e para os doentes cujo tumor apresentava ausência de expressão da proteína MSH6.

Amostra	Família	Exão /Intrão	Variante	Tipo de variante	Localização Cromossómica	Rs definido no NCBI	Frequência Alélica Europeia	Frequência Alélica Mundial	Outros softwares
L108	37	Exão 2	c.376A>T p.Thr126Ser	Missense	2:47839485	rs17036993	0,01%*	2,4%	VEP: Tolerated low confidence/benign
L1152	316	Intrão 10	c.1398+13_1398+16delATTT	Intrónica	2: 47832333-47832336	rs142115222	27%	16%	Mutation Taster: Polimorfismo
L674	56	Exão 11	c.1399-107_1399-106insAA	Frameshift	2:47823466-47823467	rs111876799	6,2%	4,1%	-
L1624	P3657	Exão 11	c.1578A>G p.Ala526=	Sinónima	2:47823181	rs35935360	0,01%*	2%	Mutation Taster: Polimorfismo; Sem alteração de local de splicing
L1121	P2209	Intrão 16	c.2083+25_2083+26insT	Intrónica	2:47813765-47813766	rs77079615	95%	91%	-
L1484 e L1855	302 e P3660	Intrão 17	c.2227+88T>G	3' UTR	2:47813146	rs3732191	7,1 %	4%	-
L5	7	Intrão 21	c.2555+53T>C	3' UTR	2:47809105	rs145978463	0,1%	0,7%	-

Legenda:As frequências alélicas foram obtidas da base de dados *Ensembl*, excepto as assinaladas* que foram obtidas da base de dados *ExAC (Exome Aggregation Consortium)*. Para as variantes que apresentavam uma baixa frequência na população procedeu-se a uma caracterização adicional recorrendo aos *softwares VEP* e/ou *Mutation Taster*, dependendo do tipo de mutação.

4.1.2. Análise de copy number para o gene *FBXO11*

A análise de variações de *copy number* no gene *FBXO11* na linha germinal de doentes de famílias FCCTX distintas foi realizada com o intuito de detectar possíveis deleções/ganhos presentes neste gene que pudessem estar associados à susceptibilidade para o FCCTX nos indivíduos analisados. Os resultados obtidos não revelaram qualquer variação de *copy number* para os indivíduos estudados.

Tendo em conta o mecanismo de interacção *antisense* descrito entre os genes *FBXO11* e *MSH6* (descrito por (Duan et al., 2012)) foram também analisadas variações de *copy number* para o gene *FBXO11* na linha germinal de doentes cujos tumores apresentavam ausência de expressão da proteína MSH6. Estes casos seriam potenciais candidatos a apresentar alterações no gene *FBXO11*. Para os indivíduos com tumores com ausência de expressão da proteína MSH6 também não foi detectada

qualquer variação de *copy-number* (figura 4.2). Os resultados obtidos para as duas *probes* utilizadas (*Probe 1*-chr 2: 26965666-2696573; *Probe 2*-chr 2: 26864659-26864712) foram coerentes entre elas, corroborando a ausência de ganhos/delecções no gene *FBXO11*.

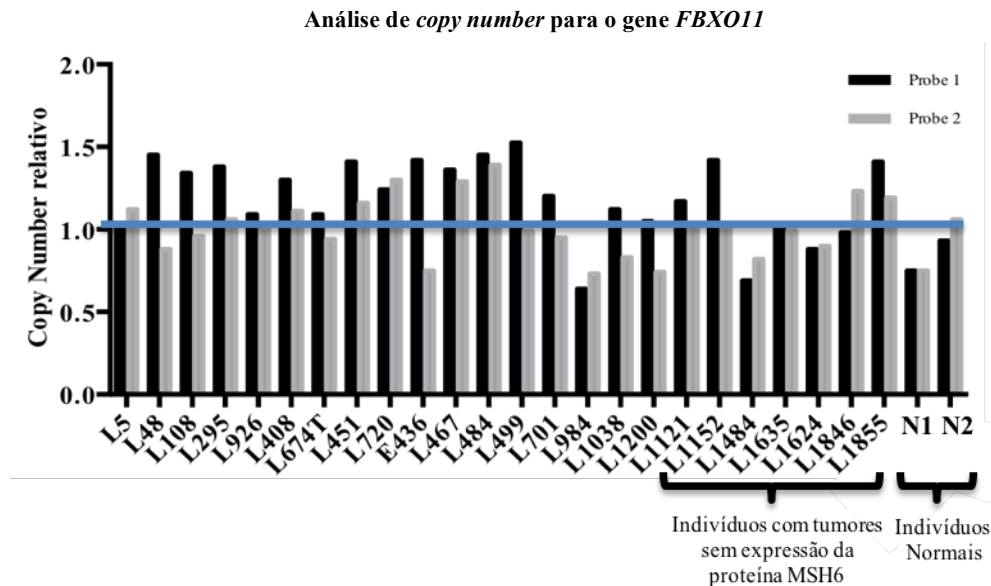


Figura 4.2. Resultados obtidos através da análise de *copy number* para o gene *FBXO11* normalizada contra o controlo *Ch12 probe*, representado pela linha a azul.

4.1.3. Análise de Expressão do gene *FBXO11*

De modo a continuar a testar a hipótese colocada, de interação entre os dois genes, gerada pela descrição do mecanismo de interação *antisense* entre os genes *FBXO11* e *MSH6* (descrito por (Duan et al., 2012) procedeu-se a uma análise de expressão do gene *FBXO11* na linha germinal de doentes cujos tumores apresentavam ausência de expressão da proteína MSH6. Devido à ausência de expressão da proteína MSH6 no tumor, estes casos seriam potenciais candidatos a apresentar alteração de expressão do gene *FBXO11*. Os resultados obtidos não indicam qualquer alteração de expressão neste gene em indivíduos com tumores sem expressão de *MSH6* (figura 4.3). Para estes casos, a ausência de expressão do gene *MSH6* não parece estar associada ao mecanismo de *antisense* entre os dois genes pois a expressão relativa do gene *FBXO11* permanece normal (intervalo de 0,5 a 1,5).

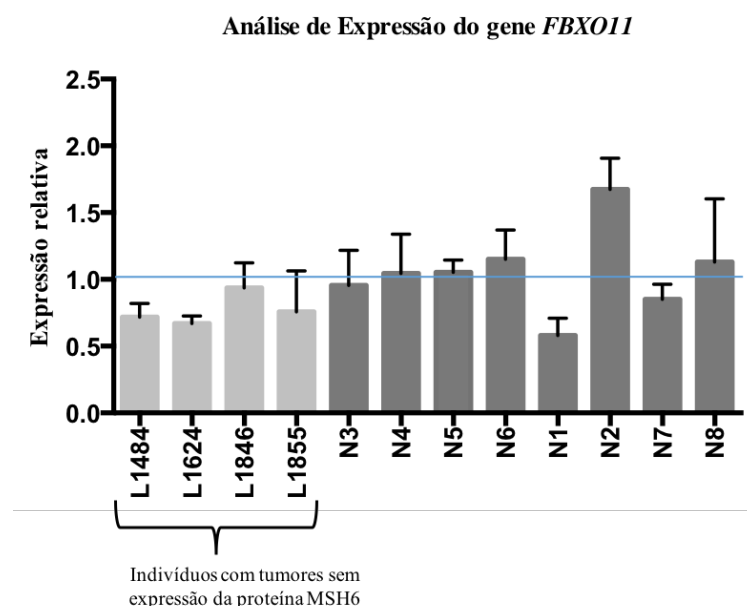


Figura 4.3. Resultados obtidos em relação à expressão relativa do gene *FBXO11*, na linha germinal de indivíduos cujos tumores apresentavam ausência de expressão da proteína MSH6 e de indivíduos normais, normalizada contra a expressão do gene controle *SDHA* (representado pela linha a azul).

4.2 Caracterização a nível germinal, das famílias FCCTX (subgrupos TSG⁺ e TSG⁻) através da análise de um painel multigénico por *Next Generation Sequencing*

A análise mutacional que envolveu 15 indivíduos index de famílias FCCTX através do painel multigénico *Trusight Cancer Panel* (Illumina) permitiu identificar em 7 famílias, todas TSG⁺, 8 mutações potencialmente patogénicas (7/15; 47% das famílias) (tabela 4.2). Estas mutações foram seleccionadas de acordo com os critérios definidos para a análise *in silico* (3.4.2.1) e apresentam todas uma elevada probabilidade de patogenicidade (geralmente *deleterius/probably damaging* nos *softwares SIFT* e *Polyphen* e um *score* elevado no *software Mutation Taster*), apresentando por vezes também alterações ao nível do *splicing*. As mutações referidas foram detectadas em genes que codificam proteínas envolvidas na reparação do DNA por recombinação homóloga (HR), nomeadamente nos genes *RECQL4* (*RecQ like helicase 4*), *CDC73* (*Cell division cycle 73*), *FANCA* (*Fanconi anemia complementation group A*), *WRN* (*Werner syndrome recQ like helicase*), *WT1* (*Wilms tumor 1*), *FANCE* (*Fanconi anemia complementation group E*), *SLX4* (*SLX4 structure-specific endonuclease subunit*) e *XPC* (*XPC complex subunit, DNA damage recognition and repair factor*). Este sistema de reparação está envolvido na reparação de quebras de dupla-cadeia (DSB) no DNA, causadas por stress genotóxico.

Apenas no indivíduo 1038, pertencente à família 173, foram identificadas duas mutações em dois genes (*FANCE* e *SLX4*), contabilizando-se para os restantes casos apenas uma. A mutação p.Arg1445Trp, identificada no gene *CDC73*, foi seleccionada apesar de classificada como *tolerated* pelo *software SIFT*, apresentando um *score* de 0,07. Tendo em conta que o *score* neste *software* varia entre 0 e 1 e um *score* inferior ou igual a 0,05 classifica uma mutação como provavelmente patogénica; tendo também em conta que é classificada pelo *software Mutation Taster* como provavelmente patogénica, decidiu-se seleccionar esta mutação como potencialmente patogénica.

Em estudos paralelos a este trabalho tinham sido também identificadas mutações possivelmente patogénicas nos indivíduos index L48, L108 e L451. Para o indivíduo L48 foi detectada uma grande deleção nos exões 5-8 do gene *SMAD4*, enquanto que no indivíduo L108 foi identificada uma mutação do tipo *missense* (p.Pro300Leu) no exão 8 do gene *POLD1* (gene que não se encontra incluído no painel *Trusight Cancer*). No doente L451, a análise de mutações nos genes que conferem predisposição para CCR e analisados em contexto de diagnóstico genético revelou a mutação p.Leu157Pro no gene *BMPRIA*. É de salientar que nenhum dos restantes indivíduos index de famílias FCCTX incluídos neste estudo revelou qualquer mutação nestes ou noutros genes que conferem predisposição para CCR e que são normalmente analisados no contexto de diagnóstico genético (*APC*, *MUTYH*, genes *MMR*, *BMPRIA*, *SMAD4*, *PTEN*, *STK11*, *TP53*).

Tabela 4.2. Descrição e caracterização das mutações germinais potencialmente patogénicas identificadas nos 15 indivíduos index de famílias FCCTX através da análise de um painel multigénico (Trusight Cancer).

Identificação da família		Identificação da amostra (indivíduo index)	Mutações provavelmente patogénicas em genes de reparação do DNA associados à HR	Frequência Mundial (EXAC)	Frequência Europeia (EXAC)	Sift/Polyphen-2	Mutation Taster
TSG ⁺	7	L5	<i>RECQL4</i> : c.2761G>A, p.Lys323Arg	0,006%	0,003%	<i>Deleterious/Possibly damaging</i>	Não foi possível obter informações do Mutation Taster para esta variante
	24	L48 [‡]	<i>CDC73</i> : c.691T>C, p.Arg1445Trp	-	-	<i>Tolerated/Possibly damaging</i>	<i>Disease causing</i> (Score: 101)
	37	L108*	<i>FANCA</i> : c.1007-3C>A	-	-	-	<i>Disease causing</i> Perda do local aceitador de <i>splicing</i>
	55	L295	-	-	-	-	-
	63	L451 [#]	-	-	-	-	-
	71	L467	-	-	-	-	-
	84	L484	<i>WRN</i> : c.435G>T, p.Trp145Cys	-	-	<i>Deleterious/Probably damaging</i>	<i>Disease causing</i> (Score:215)
	148	L984	<i>WT1</i> : c.1109G>C, p.Arg370Pro	0,0008 %	-	<i>Deleterious/Probably damaging</i>	<i>Disease causing</i> (Score:103)
	173	L1038	<i>FANCE</i> : c.1316G>T, p.Gly439Val	-	-	<i>Deleterious/Probably damaging</i>	<i>Disease causing</i> (Score:109) Perda do local dador de <i>splicing</i>
<i>SLX4</i> : c.4301C>T, p.Ser1434Leu			-	0,002%	<i>Deleterious/Probably damaging</i>	<i>Disease causing</i> (Score: 145)	
TSG ⁻	236	L1200	<i>XPC</i> : c.860T>G, p.Phe287Cys	0,2%	0,0015%	<i>Deleterious/Probably damaging</i>	<i>Disease causing</i> (Score:205)
	56	L408	-				
	69	E436	-				
	86	L499	-				
	106	L701	-				
	P2209	L1121	-				

Legenda: HR- Recombinação homóloga. Em 3 doentes assinalados foram também detectadas em estudos paralelos a este trabalho mutações potencialmente patogénicas, nos genes *SMAD4*, *POLD1* e *BMPRIA*: ‡ - grande deleção nos exões 5-8 do gene *SMAD4*; *- mutação no gene *POLD1* (p.Pro300Leu) ; # mutação no exão 10 do gene *BMPRIA* (p.Leu157Pro). No indivíduo L484 e respectiva família (L84), já tinha sido identificada uma mutação do tipo *missense* no gene *APC* (p.Glu1317Gln) num estudo prévio (Francisco et al, 2011).

Todas as mutações potencialmente patogénicas, identificadas em genes que codificam para proteínas associadas à reparação de DNA por recombinação homóloga (HR), foram detectadas em famílias classificadas como TSG⁺, cujos tumores apresentam uma assinatura molecular característica (frequentes mutações somáticas no gene *KRAS* e/ou metilação nos genes *MGMT*/MMR (tabela 4.3.). É de notar que não foram identificadas mutações potencialmente patogénicas nos indivíduos de famílias FCCTX cujos tumores não apresentavam esta assinatura molecular.

Tabela 4.3. Descrição dos tumores apresentados nas famílias FCCTX estudadas e classificação clínica e molecular dos tumores dos 15 indivíduos index de famílias FCCTX, estudados através do painel multigénico.

Identificação da Família		Caracterização clínica das famílias (número de tumores)	Caracterização somática					
			Tumor	Localização tumoral	Mutações somáticas pontuais/ perda no gene <i>APC</i>	Mutações somáticas no gene <i>KRAS</i>	Metilação no gene <i>MGMT</i>	Metilação nos genes MMR
TSG ⁺	7	CCR (3); CM (1); AT (5)	437	recto	LOH*	p.Gly12Ser	M	NM
	24	CCR (4); CG (2); CT (1); AT (3);	369	proximal	LOH	p.Gly12Ala	NM	M
	37	CCR (3)	555	proximal	p.Glu1374Lys / LOH	p.Gly12Val	NM	NM
	55	CCR (1); AT (15); ATV (4)	991	recto	p.Leu1489fs / LOH; p.Glu1353Asp; p.Gln1429His	p.Gly12Val	NM	M
	63	CCR (3); CG (1); AT (5); ATV (2)	595	distal	p.Gln1429X / LOH	N	M	M
	71	CCR (3); CG (1); AT (6)	519	distal	p.Asn1473Thr	N	NE	NE
	84	CCR (3); CT (1); AT (5)	627	recto	p.Glu1309fs / LOH	N	NM	NM
	148	CCR (3); AT (8)	1277	recto	LOH	N	M	NM
	173	CCR (3); CM (1); AT (2); AS (3)	1235	recto	N	p.Gly12Asp	NM	NM
	236	CCR (5); AT (3)	1613	recto	p.Gln1303X / LOH	N	NM	NM
TSG ⁻	56	CCR (4); CM (1); AT (14)	995	proximal	N	N	M	NM
	69	CCR (1); CG (3); CM (1); AT (1)	E436T	recto	N	N	NM	NM
	86	CCR(3); AT (6); AS (3)	741	recto	p.Glu1408X	p.Gly12Ser	NM	NM
	106	CCR (3)	L701T	recto	p.Arg1336X; p.Asn1455fs	N	NM	NM
	P2209	CCR (3); AT(1)	1541	proximal	N	p.Gly12Ala	NM	NM

Legenda:- Estes resultados foram publicados em Fam Cancer. 2011 Dec;10(4):623-31; HR - recombinação homóloga; CCR - carcinoma colorectal; CM- carcinoma da mama; AT -adenoma tubular; LOH – perda de heterozigotia (loss of heterozygosity); *- a análise de mutações somáticas não foi efectuada devido à falta de material biológico; M - metilação; NM - não metilado; CG - carcinoma gástrico; CT - carcinoma da tiróide; ATV - adenoma tubulo -viloso; N- Normal; NE - não efectuada devido à falta de material biológico; AS - adenoma serreado

Tendo em conta as mutações identificadas em genes de reparação associados à HR nos indivíduos de famílias FCCTX estudados, colocou-se a hipótese de indivíduos índice de famílias HNPCC e/ou indivíduos que preenchem os critérios de Bethesda (CB), nomeadamente, os critérios associados a história familiar de CCR incompleta (critérios 4 e 5), poderem também apresentar mutações em genes de reparação do DNA, associados à HR ou outros, o que reforçaria a associação de mutações nesses genes ao risco aumentado de CCR. Assim, dos genes incluídos no painel multigénico foram seleccionados apenas os intervenientes em sistemas de reparação de DNA das diferentes vias (BER, HR, NER (*nucleotide excision repair*) e NHEJ (*Non-homologous end joining*)) para serem avaliados num grupo de 43 doentes com HNPCC e/ou que preenchem os CB (35). Estes 78 doentes foram seleccionados por se tratarem de potenciais casos FCCTX, tendo por base o fenótipo dos mesmos ou das respectivas famílias e ausência de mutações patogénicas nos genes MMR e *MUTYH*.

A análise apenas dos genes de reparação do DNA do painel revelou que à semelhança dos indivíduos de famílias FCCTX, a maioria das mutações potencialmente patogénicas nestes doentes estavam associadas a genes de reparação do DNA associados à HR. Foi possível identificar mutações potencialmente patogénicas em genes associados à reparação por HR em 16/43 (37%) dos doentes HNPCC e em 13/35 (37%) dos doentes que preenchem os CB. Foram identificadas ainda mutações noutros genes envolvidos na reparação do DNA em 5/43 (12%) e em 3/35 (9%) destes doentes, respectivamente (tabela 4.4).

Tabela 4.4. Resultados da análise mutacional dos genes de reparação do DNA do painel em doentes de famílias HNPCC e em doentes que preenchem os CB.

		Frequência de doentes com mutações provavelmente patogénicas nos genes de reparação do DNA (%)	Genes (número de mutações únicas detectadas)
Doentes com HNPCC (n=43)	Genes de reparação do DNA associados à HR	16/43 (37%)	<i>RAD51D</i> (2), <i>ATM</i> (4), <i>FANCM</i> (2), <i>FANCC</i> (1), <i>FANCB</i> (2), <i>BLM</i> (1), <i>SLX4</i> (4), <i>XPC</i> (1)
	Outros genes de reparação do DNA	5/43 (12%)	<i>DICER1</i> (1), <i>ERCC2</i> (1), <i>ERCC3</i> (1), <i>ERCC4</i> (1), <i>TSC2</i> (1)
Doentes que preenchem os CB (n=35)	Genes de reparação do DNA associados à HR	13/35 (37%)	<i>WRN</i> (2), <i>FANCC</i> (2), <i>ATM</i> (3), <i>SLX4</i> (2), <i>BRIPI</i> (1), <i>FANCM</i> (3), <i>BLM</i> (1), <i>RAD51D</i> (1)
	Outros genes de reparação do DNA	3/35 (9%)	<i>ERCC2</i> (1), <i>ERCC3</i> (1), <i>TSC2</i> (1)

4.3 Identificação de novos genes candidatos para a susceptibilidade para o FCCTX através de sequenciação do exoma

4.3.1 Selecção de variantes identificadas através da sequenciação do exoma

A análise bioinformática efectuada após o estudo de sequenciação do exoma (WES) da família L56 foi realizada com a finalidade de seleccionar variantes germinais que pudessem esclarecer a susceptibilidade da família para FCCTX. Após a aplicação de critérios específicos para a selecção de variantes a partir da análise bioinformática (descritos no 3.2.5.1), obtiveram-se 265 variantes para o *output* 3 e 31 variantes para o *output* 4, resultando um total de 296 variantes (tabela 4.5). Uma completa caracterização, avaliação da frequência das variantes e uma análise *in silico* mais completa (3.2.5.1) permitiram restringir as 296 variantes para apenas 41, 15 do *output* 4 enquanto que as restantes 26 do *output* 3 (tabela 4.5).

Foram ainda contabilizadas mais 3 variantes provenientes de uma análise bioinformática preliminar realizada anteriormente na família, resultando 44 variantes que prosseguiram para o estudo de segregação na família 56 (tabela 4.6).

Tabela 4.5. Resumo do número de variantes iniciais obtidas pela análise bioinformática, após a aplicação dos critérios de selecção e análise *in silico* nas diferentes análises e *outputs*.

		Análise 1	Análise 2	Análise 3	Análise 4	Análise 5	Total de variantes
Output 3	Nº de variantes iniciais obtidas pela análise bioinformática	12 603	2 192	1 422	890	1 535	18 642
	Nº de variantes obtidas após a aplicação dos critérios de selecção	*	74	55	46	90	265
	Nº de variantes obtidas após análise <i>in silico</i>	*	4	8	4	10	26
Output 4	Nº de variantes iniciais obtidas pela análise bioinformática	353	89	62	29	71	604
	Nº de variantes obtidas após a aplicação dos critérios de selecção	4	7	5	7	8	31
	Nº de variantes obtidas após análise <i>in silico</i>	3	1	-	5	6	15
Número total de variantes para análise de segregação		3	5	8	9	16	41

Legenda: * a análise 1 para o *output* 3 não foi realizada na sua totalidade por indisponibilidade de tempo.

Tabela 4.6. Caracterização das variantes seleccionadas após a aplicação dos critérios de selecção e critérios mais restritos em termos de frequência das mutações e de análise *in silico* (ponto 3.2.5.1).

Output	Análise	Gene	Variante em estudo	Tipo de variante	Rs definido NCBI	Frequência Alélica Mundial (%)		Frequência Alélica Europeia (%)		Previsão <i>In silico</i>	
						Ensembl	Exac	Ensembl	Exac	<i>Sift/Polyphen</i>	<i>Mutation taster</i>
Output 4	1	<i>MAML2</i>	c.1821_1829delACAGCAGCA;p.Gln472_Gln474del	<i>Inframe deletion</i>	rs750046576	0,001	-	-	-	-	-
		<i>MTMR3</i>	c.1933C>T;p.Arg645Trp	<i>Missense</i>	rs138823197	0,02	0,020	-	0,03	<i>Deleterious/probably damaging</i>	<i>Disease Causing (Score:101)</i>
		<i>RNF207</i>	c.1615C>T;p.Arg2695Cys	<i>Missense</i>	rs55823245	0,90	1,50	2,6	2,10	<i>Deleterious/possibly damaging</i>	<i>Disease Causing (Score:180)</i>
	2	<i>OR3A1</i>	c.392C>T;p.Arg131Gln	<i>Missense</i>	rs201927505	0,03	0,03	-	0,01	<i>Tolerated/benign</i>	Polimorfismo
		<i>C2orf70</i>	c.232C>G;p.Pro78Ala	<i>Missense</i>	rs112873448	0,7	1,80	2,50	2,70	<i>Deleterious/probably damaging</i>	Polimorfismo
	4	<i>GP1BA</i>	c.775G>A;p.Val259Met	<i>Missense</i>	rs749788312	0,03	-	-	-	<i>Deleterious/probably damaging</i>	Polimorfismo
		<i>MAG11</i>	c.1263_1264insCAG p. Gln421_Thr422insGln	<i>Frameshift</i>	rs142043619	0,30	2,60	-	2,40	-	-
		<i>NPR2</i>	c.2644G>A;p.Val882Ile	<i>Missense</i>	rs55700371	0,01	-	0,01	0,01	<i>Deleterious/possibly damaging</i>	<i>Disease Causing (Score:29)</i>
		<i>PROSER2</i>	c.607G>A;p.Glu203Lys	<i>Missense</i>	rs2147439	0,30	1,40	1	2,10	<i>Deleterious/probably damaging</i>	<i>Disease Causing (Score: 56)</i>
		<i>C16orf96</i>	c.611C>T;p.Ser204Phe	<i>Missense</i>	rs62037460	0,01	0,01	-	0,02	<i>Deleterious/probably damaging</i>	Polimorfismo
	5	<i>DCHS1</i>	p.Cys34_Cys35ins	<i>Frameshift</i>	-	-	-	-	-	-	-
		<i>DUSP12</i>	c.6_10insTTG; p.Leu2_Leu3ins;	<i>Frameshift</i>	-	-	-	-	-	-	<i>Disease Causing (Score:138)</i>
		<i>OR56A4</i>	c.464T>C; p.Arg155Gln	<i>Missense</i>	-	-	-	-	-	-	<i>Disease Causing (Score:76)</i>
		<i>SMG7</i>	c.1970C>A; p.Pro657His	<i>Missense</i>	rs34426362	0,20	0,40	0,70	0,60	<i>Deleterious/probably damaging</i>	<i>Disease Causing (Score: 77)</i>
		<i>TNIP3</i>	c.854T>G; p.Glu285Ala	<i>Missense</i>	rs75386686	1,30	2,80	2,70	2,70	<i>Deleterious/probably damaging</i>	Polimorfismo
		<i>GALNT8</i>	c.988G>A;p.Glu330Lys	<i>Missense</i>	rs201387598	0,01	0,01	0,10	-	<i>Tolerated/benign-</i>	Polimorfismo
Output 3	2	<i>MAP4</i>	c.6422G>A; p.Ser2141Phe	<i>Missense</i>	rs150907099	0,20	0,30	0,30	0,40	<i>Deleterious/probably damaging</i>	Polimorfismo
		<i>PTPN23</i>	c.3988G>A;p.Val1330Met	<i>Missense</i>	rs141816989	0,10	0,08	0,10	0,08	<i>Deleterious/probably damaging</i>	<i>Disease Causing (Score:21)</i>
		<i>SULT1A1</i>	c..44C>A	<i>5'UTR</i>	rs143643294	0,30	0,40	0,80	0,60	-	-
		<i>ARHGAP22</i>	c.963G>A; p.Asn321=	<i>Synonimous</i>	rs199896839	0,01	0,01	-	0,01	-	<i>Disease Causing; Ganho do local dador de splicing (0.71)</i>

Continuação da tabela 4.6. Caracterização das variantes seleccionadas após a aplicação dos critérios de selecção e critérios mais restritos em termos de frequência das mutações e de análise *in silico* (ponto 3.2.5.1).

Output	Análise	Gene	Variante em estudo	Tipo	Rs definido NCBI	Frequência Alélica Mundial (%)		Frequência Alélica Europeia (%)		Previsão <i>In silico</i>	
						Ensembl	Exac	Ensembl	Exac	Sift/Polyphen	Mutation taster
Output 3	3	CACNA1S	c.3607C>T; p.Asp1203Asn	Missense/Splicing	rs367956917	0,002	0,002	-	-	Tolerated/possibly damaging	Disease Causing (Score:160);Perda do dador de splicing;
		KIAA1462	c.1660T>C; p.Thr554Ala	Missense	rs201339774	0,03	0,03	-	0,04	Tolerated/benign	Disease causing (Score:38)
		MAP6	c.1716A>G; p.Asn572=	Synonymous	rs61739609	0,01	0,08	-	0,10	-	Polimorfismo; Ganho de local dador de splicing (0.95)
		NEK7	c.103C>G; p.Arg35Gly	Missense	rs55833332	0,40	0,60	0,90	0,80	Tolerated/benign	Disease Causing (Score:125)
		PARD3	c.2540G>A; p.Ala847Val	Missense	rs16935326	0,40	0,30	0,30	0,20	Tolerated/benign	Disease Causing (Score:78)
		RASL11B	c.577G>A; p.Ala193Thr	Missense	rs28416473	0,70	0,20	-	0,05	Deleterious/possibly	Disease Causing (Score:58)
		WDR37	c.531C>T; p.Ala177=	Splicing	rs145160560	0,10	0,20	0,50	0,30	-	Disease Causing; Perda do dador de splicing;
	4	LGR6	c.1868G>A; p.Arg623Gln	Missense	rs767586087	0,002	-	-	-	Tolerated/benign	Disease Causing (Score:126)
		PELP1	c.894C>T; p.Ser298=	Synonymous	rs368343383	0,04	0,003	-	0,003	-	Disease Causing; Sem efeito a nível de splicing;
		RNF217	c.6+3A>G	Splicing	rs139617496	0,20	0,08	-	-	-	-
		TM4SF5	c.179T>C; p.Val60Ala	Missense	rs757673757	0,002	-	-	-	Deleterious/benign	Disease Causing (Score:64)
	5	ANKRD12	c.297A>G; p.Thr99=	Synonymous	rs112568144	0,40	0,60	0,50	0,50	-	Disease Causing;Ganho de local dador de splicing (0.94)
		AKR1C3	c.623C>T; p.Ser208Leu	Missense	rs116351688	0,10	0,001	-	0,003	Deleterious/benign	Polimorfismo
		BRD4	c.3810C>G; p.Glu1270Asp	Missense	rs113881069	0,10	0,50	0,30	0,70	Tolerated/benign	Disease Causing (Score:45)
		BBOX1	c.639+5_639+6insG	Splicing	rs397839995	-	0,0001	-	0,001	-	-
		CEP250	c.4208A>C; p.Gln1403Pro	Missense	-	-	-	-	-	Deleterious/probably damaging	Polimorfismo
		CLIP2	c.2238G>A; p.Ala746=	Synonymous	rs146488093	0,10	0,03	0,10	0,04	-	Disease Causing; Ganho de local dador de splicing (0.83)
		PFKP	c.339C>T; p.Asn113=	Splicing	rs201731667	0,01	0,01	-	0,01	-	Disease Causing
		PTPN22	c.197-4A>G	Splicing	rs72650669	0,10	0,20	0,20	0,30	-	-
		TP63	c.-72 -71insAGAG	5'UTR	rs34201045	-	-	-	-	-	-
		ZNF75A	c.931A>G	Splicing	-	-	-	-	-	-	-

4.3.2 Estudo de segregação das variantes seleccionadas, identificadas através da sequenciação do exoma

O estudo de segregação com a doença na família das variantes seleccionadas, identificadas pela sequenciação do exoma, foi realizado através da avaliação da presença/ausência da variante nas respectivas amostras de indivíduos afectados e não afectados pertencentes à família em análise. Esta informação juntamente com o fenótipo dos indivíduos permitiu concluir sobre a segregação da mutação ou não na família. A análise bioinformática tinha sido dividida em 5 sub-análises tendo em conta a exclusão de indivíduos que poderiam eventualmente ser possíveis fenocópias e assim seleccionar variantes mais fidedignas num contexto de síndrome hereditária.

A análise 1 poderá ser a análise mais informativa na medida que retrata as variantes partilhadas pelos indivíduos afectados sem a possibilidade de haver casos de origem esporádica. Das 3 variantes pertencentes a esta análise apenas a variante *missense* p.Arg645Trp do gene *MTMR3* (22q12.2) segregou totalmente com a doença na família (tabela 4.7).

Tabela 4.7. Resultados da análise de segregação das variantes seleccionadas a partir dos resultados de *output* da análise1. Na tabela estão descritos os genes onde se identificaram as variantes. Para cada indivíduo foi realizada uma breve caracterização clínica, incluindo o número de carcinomas e de pólipos.

Amostra (Família L56)	Informação clínica (idade de diagnóstico)	Análise 1		
		<i>MAML2</i>	<i>MTMR3</i>	<i>RNF207</i>
L408 (Indivíduo 3)	1CA(45A), 3Ad(66A,71A,75A), 3PH(71A,72A,75A)	Het	Het	Het
L674 (Indivíduo 1)	1 CA(34A), 1Ad(36A), 3PH(37A,38A,38A)	Normal	Het	Het
L676	2 Ad (43A,44A), 1PH (41A)	Normal	Het	Het
L781 (Indivíduo5)	8Ad(2(32A),1(37A),1(38A),2(42 A),2(43A)), 4 PH(2(32A),2(43A))	Het	Het	Het
L801 (Indivíduo 2)	1 CA (57A), 1 Ad (57A), 2PH(64A,65A)	Het	Het	Het
L802	Não afectado	Het	Normal	Normal
L860	Não afectado	Het	Normal	Normal
L1094	2 Ad (47A,51A)	Het	Normal	Het
L1104	5 Ad (71A) ,1Ad(72A)	Het	Het	Het
L1793	1 Ad e 3 ASS(*), 1 Ad (55A)	Normal	Het	Normal
L1794 (Indivíduo 4)	2 Ad(47A,51A), 1PH(40A)	Het	Het	Het
L1956	Não afectado	Normal	Normal	Het

Legenda: CA-carcinoma, Ad-adenoma, PH-pólipo hiperplásico, ASS- adenoma serrado sésil. Cada amostra correspondente ao indivíduo da família em estudo que foi identificada como: Normal -quando não possuía a variante; Het- variante presente em heterozigotia; Homo- variante presente em homozigotia; Os indivíduos assinalados a cores são os indivíduos integrados na análise 1, ou seja as variantes identificadas na análise 1 são partilhadas por esses indivíduos. A idade de diagnóstico (em anos, A) e respectiva caracterização fenotípica encontra-se descrita para os indivíduos afectados. Todos os adenomas identificados correspondem a adenomas tubulares com displasia de baixo grau, com a excepção no indivíduo L1104 que apresenta um adenoma tubular com displasia de alto grau aos 71A. (*) Não foi possível obter-se idade de diagnóstico.

Para além da presença da mutação em todos os indivíduos para os quais foi efectuada a sequenciação do exoma, os quais foram todos incluídos neste modelo de análise 1 (indivíduos 1-5), a alteração p.Arg645Trp estava ausente em indivíduos saudáveis e presente nos indivíduos afectados (figura 4.4), excepto no indivíduo L1094. Esta mutação foi considerada como tendo segregado na

família visto que este indivíduo desenvolveu até ao momento apenas 2 adenomas, o que não se pode excluir que possa ter acontecido no contexto esporádico ou eventualmente de outra variante.

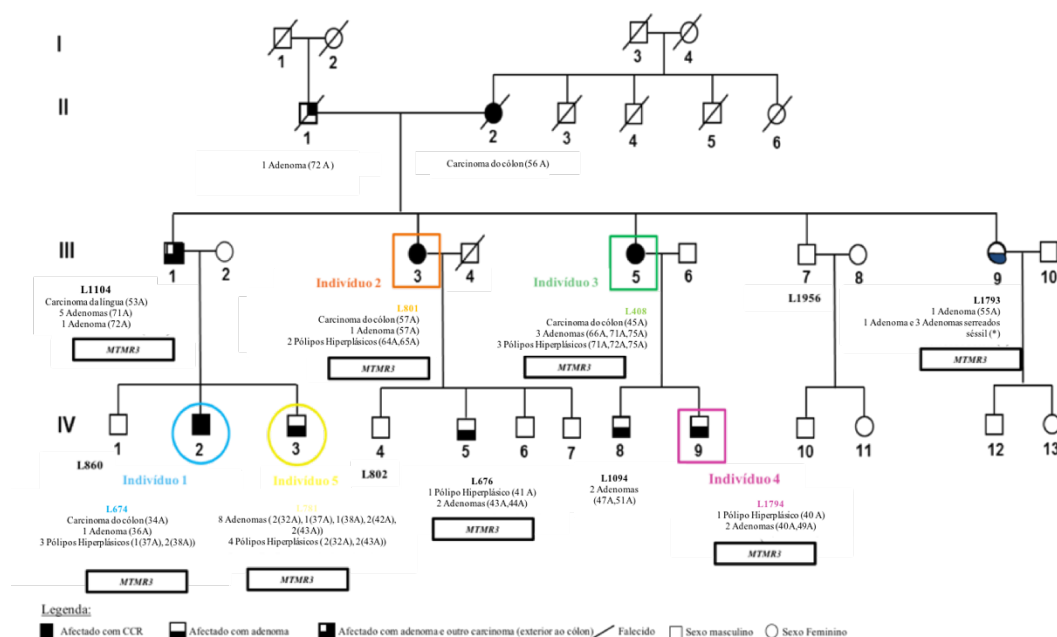


Figura 4.4. Os indivíduos assinalados a cores foram seleccionados para análise bioinformática, no que diz respeito à análise 1. Apenas para os indivíduos com número de identificação LXXX ou LXXXX existe material biológico disponível. Para os indivíduos afectados foi feita uma caracterização fenotípica e respectiva idade de diagnóstico com a excepção de * para o qual não se dispunha de informação disponível. Todos os adenomas identificados correspondem a adenomas tubulares com displasia de baixo grau, com a excepção no indivíduo L1104 que apresenta um adenoma tubular com displasia de alto grau aos 71A.

A análise 2 abrange as variantes partilhadas por quatro indivíduos (1, 2, 3 e 5) e a exclusão do indivíduo 4 visto este ter desenvolvido apenas 2 adenomas até aos 51 e ainda não ter desenvolvido carcinoma, podendo haver a possibilidade de ser um caso de origem esporádica. Não se observou a segregação de nenhuma variante proveniente desta análise visto que estavam presentes tanto em indivíduos saudáveis como em indivíduos afectados (tabela 4.8).

Tabela 4.8. Resultados da análise de segregação das variantes seleccionadas a partir dos resultados de *output* da análise 2. Na tabela estão descritos os genes onde se identificaram as variantes. Para cada indivíduo foi realizada uma breve caracterização clínica, incluindo o número de carcinomas e de pólipos.

Amostra (Família L56)	Informação clínica (idade de diagnóstico)	Análise 2				
		<i>GALNT8</i>	<i>MAP4</i>	<i>OR3A1</i>	<i>PTPN23</i>	<i>SULT1A1</i>
L408 (Indivíduo 3)	1CA(45A), 3Ad(66A,71A,75A), 3PH(71A,72A,75A)	Het	Het	Het	Het	Het
L674 (Indivíduo 1)	1 CA(34A), 1Ad(36A), 3PH(37A,38A,38A)	Het	Het	Het	Het	Het
L676	2 Ad (43A,44A), 1PH (41A)	Het	Het	Het	Het	Normal
L781 (Indivíduo5)	8Ad(2(32A),1(37A),1(38A),2(42A),2(43A)), 4 PH(2(32A),2(43A))	Het	Het	Het	Het	Het
L801	1 CA (57A), 1 Ad (57A), 2PH(64A,65A)	Het	Het	Het	Het	Het

(Indivíduo 2)						
L802	Não afectado	Normal	Het	Het	Het	Normal
L860	Não afectado	Het	Het	Normal	Het	Het
L1094	2 Ad (47A,51A)	Normal	Normal	Normal	Normal	Het
L1104	5 Ad (71A), 1Ad(72A)	Het	Het	Het	Het	Het
L1793	1 Ad e 3 ASS (*), 1 Ad (55A)	Normal	Normal	Homo	Normal	Het
L1794 (Indivíduo 4)	2 Ad(47A,51A), 1PH(40A)	Het	Normal	Normal	Normal	Normal
L1956	Não afectado	Het	Normal	Het	Normal	Het

Legenda: CA-carcinoma, Ad-adenoma, PH-pólipo pólo hiperplásico, ASS- adenoma serreado sésil. Cada amostra correspondente ao indivíduo da família em estudo que foi identificada como: Normal -quando não possuía a variante; Het- variante presente em heterozigotia; Homo- variante presente em homozigotia; Os indivíduos assinalados a cores são os indivíduos integrados na análise 2, ou seja as variantes identificadas na análise 2 com a exclusão das variantes partilhadas também pelo indivíduo cinzento. A idade de diagnóstico (em anos, A) e respectiva caracterização fenotípica encontra-se descrita para os indivíduos afectados. Todos os adenomas identificados correspondem a adenomas tubulares com displasia de baixo grau, com a excepção no indivíduo L1104 que apresenta um adenoma tubular com displasia de alto grau aos 71A. (*) Não foi possível obter-se idade de diagnóstico.

A análise 3 incluiu as variantes partilhadas pelos indivíduos 1, 3, 4 e 5, eliminando o indivíduo 2 da análise, ou seja considerando-o como uma eventual fenocópia. Nesta análise identificou-se a segregação parcial da variante p.Arg35Gly correspondente ao gene *NEK7* (*NIMA related kinase 7*)(1q31.3) (tabela 4.9).

Tabela 4.9. Resultados da análise de segregação das variantes seleccionadas a partir dos resultados de *output* da análise 3. Na tabela estão descritos os genes onde se identificaram as variantes. Para cada indivíduo foi realizada uma breve caracterização clínica, incluindo o número de carcinomas e de pólipos.

Amostra (Família L56)	Informação clínica (idade de diagnóstico)	Análise 3						
		<i>ARHGAP22</i>	<i>KIAA1462</i>	<i>MAP6</i>	<i>NEK7</i>	<i>PARD3</i>	<i>RASL11B</i>	<i>WDR37</i>
L408 (Indivíduo 3)	1CA(45A), 3Ad(66A,71A,75A), 3PH(71A,72A,75A)	Het	Het	Het	Het	Het	Het	Het
L674 (Indivíduo 1)	1 CA(34A), 1Ad(36A), 3PH(37A,38A,38A)	Het	Het	Het	Het	Het	Het	Het
L676	2 Ad (43A,44A), 1PH (41A)	Normal	Normal	Normal	Normal	Normal	Normal	Normal
L781 (Indivíduo 5)	8Ad(2(32A),1(37A),1(38A),2(42A),2(43A)), 4 PH(2(32A),2(43A))	Het	Het	Het	Het	Het	Het	Het
L801 (Indivíduo 2)	1 CA (57A), 1 Ad (57A), 2PH(64A,65A)	Normal	Normal	Normal	Normal	Normal	Normal	Normal
L802	Não afectado	Normal	Normal	Normal	Normal	Normal	Normal	Normal
L860	Não afectado	Het	Het	Normal	Normal	Het	Normal	Het
L1094	2 Ad (47A,51A)	Het	Het	Het	Het	Het	Het	Het
L1104	5 Ad (71A), 1Ad(72A)	Het	Het	Normal	Het	Het	Het	Het
L1793	1 Ad e 3 ASS (*), 1 Ad (55A)	Normal	Normal	Het	Het	Normal	Het	Het
L1794 (Indivíduo 4)	2 Ad(47A,51A), 1PH(40A)	Het	Het	Het	Het	Het	Het	Het
L1956	Não afectado	Normal	Normal	Het	Normal	Normal	Het	Het

Legenda: CA-carcinoma, Ad-adenoma, PH-pólipo pólo hiperplásico, ASS- adenoma serreado sésil. Cada amostra correspondente ao indivíduo da família em estudo que foi identificada como: Normal -quando não possuía a variante; Het- variante presente em heterozigotia; Homo- variante presente em homozigotia; Os indivíduos assinalados a cores são os indivíduos integrados na análise 3, ou seja as variantes identificadas na análise 3 com a exclusão das variantes partilhadas também pelo indivíduo cinzento. A idade de diagnóstico (em anos, A) e respectiva caracterização fenotípica encontra-se descrita para os indivíduos afectados. Todos os adenomas identificados correspondem a adenomas tubulares com displasia de baixo grau, com a excepção no indivíduo L1104 que apresenta um adenoma tubular com displasia de alto grau aos 71A. (*) Não foi possível obter-se idade de diagnóstico.

A maioria dos indivíduos afectados apresentava a variante p.Arg35Gly enquanto que esta não foi observada para nenhum indivíduo saudável. Para o indivíduo L801, dado não ter sido incluído na análise 3, já era esperado a variante não estar presente. Se este indivíduo não partilhar as variantes

herdadas por se tratar de um caso de origem esporádica, o seu filho (L676) de acordo com o modelo de segregação também não as irá apresentar. De facto, esta variante não foi identificada nestes dois indivíduos (figura 4.5).

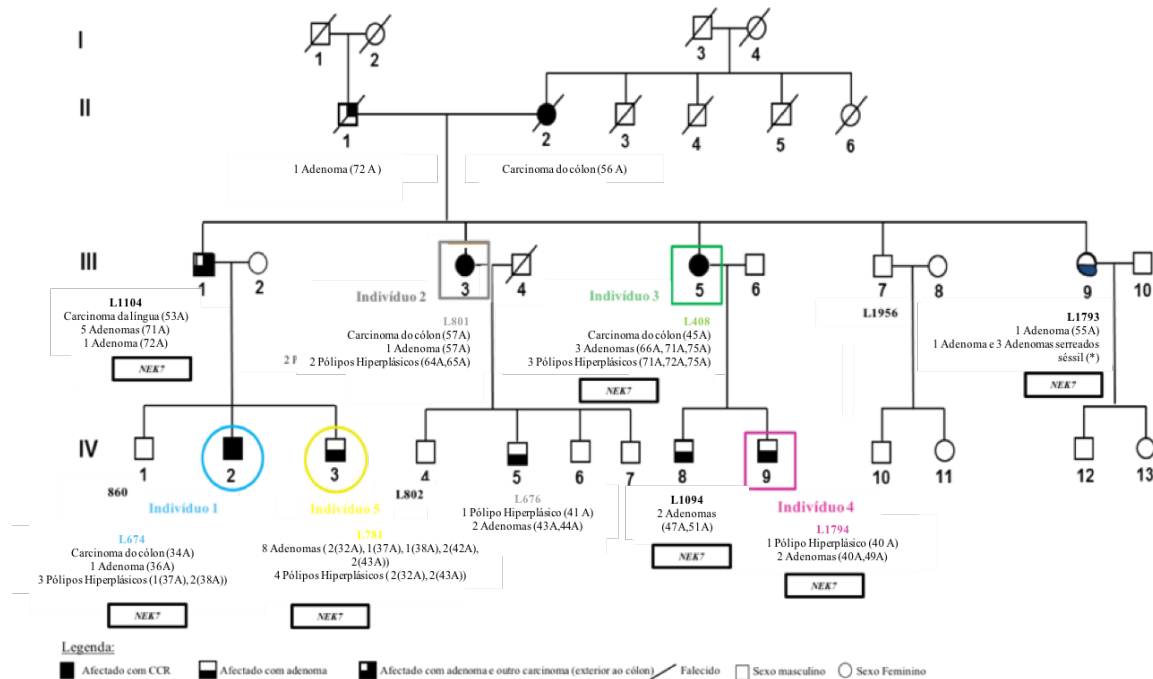


Figura 4.5. Os indivíduos assinalados a cores foram seleccionados para análise bioinformática, no que diz respeito à análise 3. Apenas para os indivíduos com número de identificação LXXX ou LXXXX existe material biológico disponível. Para os indivíduos afectados foi feita uma caracterização fenotípica e respectiva idade de diagnóstico com a excepção de * para o qual não se dispunha de informação disponível. Todos os adenomas identificados correspondem a adenomas tubulares com displasia de baixo grau, com a excepção no indivíduo L1104 que apresenta um adenoma tubular com displasia de alto grau aos 71A.

A análise 4 integrou as variantes partilhadas pelos indivíduos 1, 3 e 5, e considerou dois indivíduos como fenocópias (indivíduos 2 e 4). Desta análise resultaram 4 variantes (p.Pro78Ala, p.Val882Ile, p.Arg623Gln, p.Ser298=) que segregaram parcialmente com a doença na família, correspondendo aos genes *C2orf70* (Chromosome 2 open reading frame 70) (2p23.3), *NPR2* (Natriuretic peptide receptor 2) (9p13.3), *LGR6* (Leucine rich repeat containing G protein-coupled receptor 6) (1q32.1) e *PELP1* (Proline, glutamate and leucine rich protein 1) (17p13.2), respectivamente (tabela 4.10).

Tabela 4.10. Resultados da análise de segregação das variantes seleccionadas a partir dos resultados de *output* da análise 4. Na tabela estão descritos os genes onde se identificaram as variantes. Para cada indivíduo foi realizada uma breve caracterização clínica, incluindo o número de carcinomas e de pólipos.

	Informação clínica (Idade de diagnóstico)	Análise 4								
		<i>C2orf70</i>	<i>GPB1A</i>	<i>LGR6</i>	<i>MAGI1</i>	<i>NPR2</i>	<i>PELPI</i>	<i>PROSER2</i>	<i>RNF217</i>	<i>TM4SF5</i>
L408 (Indivíduo 3)	1CA(45A), 3Ad(66A,71A,75A), 3PH(71A,72A,75A)	Het	Het	Het	Het	Het	Het	Het	Het	Het
L674 (Indivíduo 1)	1 CA(34A), 1Ad(36A), 3PH(37A,38A,38A)	Het	Het	Het	Normal	Het	Het	Het	Het	Het
L676	2 Ad (43A,44A), 1PH (41A)	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal
L781 (Indivíduo 5)	8Ad(2(32A),1(37A),1(38A),2(42A),2(43A)), 4 PH(2(32A),2(43A))	Het	Het	Het	Normal	Het	Het	Het	Het	Het
L801 (Indivíduo 2)	1 CA (57A), 1 Ad (57A), 2PH(64A,65A)	Normal	Normal	Normal	Het	Normal	Normal	Normal	Normal	Normal
L802	Não afectado	Normal	Normal	Normal	Het	Normal	Normal	Normal	Normal	Normal
L860	Não afectado	Normal	Normal	Normal	Het	Normal	Normal	Het	Het	Normal
L1094	2 Ad (47A,51A)	Het	Normal	Het	Normal	Normal	Normal	Het	Normal	Normal
L1104	5 Ad (71A), 1Ad(72A)	Het	Het	Het	Het	Het	Het	Het	Het	Het
L1793	1 Ad e 3 ASS (*), 1 Ad (55A)	Normal	Het	Het	Het	Normal	Het	Normal	Normal	Het
L1794 (Indivíduo 4)	2 Ad(47A,51A), 1PH(40A)	Normal	Normal	Normal	Normal	Normal	Normal	Het	Normal	Normal
L1956	Não afectado	Normal	Het	Normal	Normal	Normal	Het	Normal	Normal	Het

Legenda: CA-carcinoma, Ad-adenoma, PH-pólipo pólipos hiperplásico, ASS- adenoma serrado sésil. Cada amostra correspondente ao indivíduo da família em estudo que foi identificada como: Normal -quando não possuía a variante; Het- variante presente em heterozigotia; Homo- variante presente em homozigotia; Os indivíduos assinalados a cores são os indivíduos integrados na análise 4, ou seja as variantes identificadas na análise 4 com a exclusão das variantes partilhadas também pelos indivíduos a cinzento. A idade de diagnóstico (em anos, A) e respectiva caracterização fenotípica encontra-se descrita para os indivíduos afectados. Todos os adenomas identificados correspondem a adenomas tubulares com displasia de baixo grau, com a excepção no indivíduo L1104 que apresenta um adenoma tubular com displasia de alto grau aos 71A. (*) Não foi possível obter-se idade de diagnóstico.

Com base no modelo da análise 4, a combinação das quatro variantes dos genes *C2orf70*, *NPR2*, *LGR6* e *PELPI* foi identificada em indivíduos com fenótipo mais marcado na família (L1104, L674, L781, L408) e que desenvolveram carcinoma. O indivíduo L1094 apresentou apenas uma destas variantes, a variante identificada no gene *LGR6*, enquanto que o indivíduo L1793 apenas apresentou duas destas variantes, respectivamente nos genes *LGR6* e *PELPI*. Estas variantes não foram identificadas nos indivíduos L801 e L1794 face ao modelo de análise em que se inserem. O descendente do L801 (L676) também não apresentou qualquer variante de acordo com o modelo de segregação (imagem 4.6).

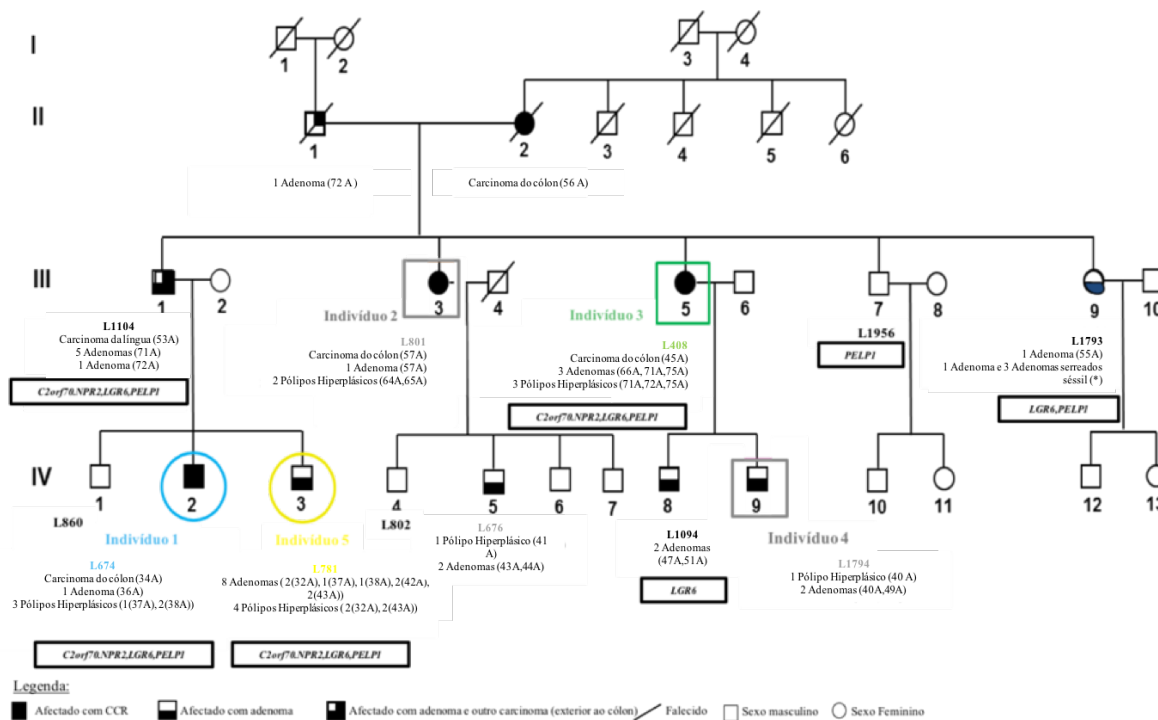


Figura 4.6. Os indivíduos assinalados a cores foram seleccionados para análise bioinformática, no que diz respeito à análise 4. Apenas para os indivíduos com número de identificação LXXX ou LXXXX existe material biológico disponível. Para os indivíduos afectados foi feita uma caracterização fenotípica e respectiva idade de diagnóstico com a excepção de * para o qual não se dispunha de informação disponível. Todos os adenomas identificados correspondem a adenomas tubulares com displasia de baixo grau, com a excepção no indivíduo L1104 que apresenta um adenoma tubular com displasia de alto grau aos 71A.

Na análise 5 foram seleccionadas as variantes partilhadas entre os indivíduos 1, 2 e 5 sendo os indivíduos 3 e 4 possivelmente fenocópias. Desta análise resultaram apenas duas variantes que segregaram parcialmente na doença com a família: p.Leu2_leu3ins e p.Pro657His, dos genes *DUSP12* (*Dual specificity phosphatase 12*) (1q23.3) e *SMG7* (*Nonsense mediated mRNA decay factor*) (1q25.3), respectivamente (tabela 4.11).

Tabela 4.11. Resultados da análise de segregação das variantes seleccionadas a partir dos resultados de *output* da análise 5. Na tabela estão descritos os genes onde se identificaram as variantes. Para cada indivíduo foi realizada uma breve caracterização clínica, incluindo o número de carcinomas e de pólipos.

Amostra (Família L56)	Informação clínica (idade de diagnóstico)	Análise 5							
		<i>ANKRD12</i>	<i>AKR1C3</i>	<i>BRD4</i>	<i>BBOX1</i>	<i>C16orf96</i>	<i>CEP250</i>	<i>CLIP2</i>	<i>DCHS1</i>
L408 (Indivíduo 3)	1CA(45A), 3Ad(66A,71A,75A), 3PH(71A,72A,75A)	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal
L674 (Indivíduo 1)	1 CA(34A), 1Ad(36A), 3PH(37A,38A,38A)	Het	Het	Het	Het	Het	Het	Het	Het
L676	2 Ad (43A,44A), 1PH (41A)	Het	Het	Het	Normal	Normal	Normal	Het	Het
L781 (Indivíduo 5)	8Ad(2(32A),1(37A),1(38A),2(42A), 2(43A)), 4 PH(2(32A),2(43A))	Het	Het	Het	Het	Het	Het	Het	Het
L801 (Indivíduo 2)	1 CA (57A), 1 Ad (57A), 2PH(64A,65A)	Het	Het	Het	Het	Het	Het	Het	Het
L802	Não afectado	Het	Het	Normal	Het	Het	Het	Normal	Het
L860	Não afectado	Homo	Het	Het	Normal	Het	Het	Normal	Normal

L1094	2 Ad (47A,51A)	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Het
L1104	5 Ad (71A), 1Ad(72A)	Het	Het	Normal	Het	Het	Het	Het	Normal
L1793	1 Ad e 3 ASS (*), 1 Ad (55A)	Normal	Normal	Normal	Normal	Het	Het	Het	Normal
L1794 (Indivíduo 4)	2 Ad(47A,51A), 1PH(40A)	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal
L1956	Não afectado	Het	Normal	Het	Normal	Normal	Het	Het	Normal
Amostra (Família L56)	Informação clínica (idade de diagnóstico)	DUSP12	OR566A4	PFKP	PTPN22	SMG7	TNIP3	TP63	ZNF75A
L408 (Indivíduo 3)	1CA(45A), 3Ad(66A,71A,75A), 3PH(71A,72A,75A)	Normal	Normal	Normal	Normal	Normal	Normal	Het	Normal
L674 (Indivíduo 1)	1 CA(34A), 1Ad(36A), 3PH(37A,38A,38A)	Het	Het	Het	Het	Het	Het	Het	Het
L676	2 Ad (43A,44A), 1PH (41A)	Het	Normal	Het	Het	Het	Normal	Het	Normal
L781 (Indivíduo 5)	8Ad(2(32A),1(37A),1(38A),2(42 A),2(43A)), 4 PH(2(32A),2(43A))	Het	Het	Het	Het	Het	Het	Het	Het
L801 (Indivíduo 2)	1 CA (57A), 1 Ad (57A), 2PH(64A,65A)	Het	Het	Het	Het	Het	Het	Het	Het
L802	Não afectado	Normal	Het	Het	Normal	Normal	Normal	Het	Het
L860	Não afectado	Normal	Normal	Het	Het	Normal	Het	Het	Het
L1094	2 Ad (47A,51A)	Normal	Normal	Normal	Normal	Normal	Normal	Het	Normal
L1104	5 Ad (71A), 1Ad(72A)	Het	Het	Het	Het	Het	Normal	Het	Het
L1793	1 Ad e 3 ASS (*), 1 Ad (55A)	Normal	Normal	Normal	Normal	Normal	Het	Het	Het
L1794 (Indivíduo 4)	2 Ad(47A,51A), 1PH(40A)	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal
L1956	Não afectado	Normal	Normal	Normal	Normal	Het	Normal	Het	Normal

Legenda: CA-carcinoma, Ad-adenoma, PH-pólipo pólo hiperplásico, ASS- adenoma serreado sésil. Cada amostra correspondente ao indivíduo da família em estudo foi identificada como: Normal -quando não possuía a variante; Het- variante presente em heterozigotia; Homo- variante presente em homozigotia; Os indivíduos assinalados a cores são os indivíduos integrados na análise 1, ou seja as variantes identificadas na análise 1 são partilhadas por esses indivíduos. A idade de diagnóstico (em anos, A) e respectiva caracterização fenotípica encontra-se descrita para os indivíduos afectados. Todos os adenomas identificados correspondem a adenomas tubulares com displasia de baixo grau, com a excepção no indivíduo L1104 que apresenta um adenoma tubular com displasia de alto grau aos 71A. (*) Não foi possível obter-se idade de diagnóstico.

Os indivíduos L1104, L674, L781, L801 e L676, apresentaram o mesmo genótipo para estas duas variantes (figura 4.7). Visto que os genes *DUSP12* e *SMG7* situam-se em regiões cromossómicas adjacentes (figura 4.9) é coerente que as variantes presentes nos mesmos tenham segregado em conjunto nos mesmos indivíduos. É ainda de salientar que estas alterações foram observadas em indivíduos com carcinoma ou com mais adenomas e em idade mais jovem. Tendo em conta o modelo da análise 5, se os indivíduos 3 (L408) e 4 (L1794) forem casos esporádicos ou associados a outra qualquer variante é de esperar que o seu familiar directo L1094 também não tenha a variante germinal.

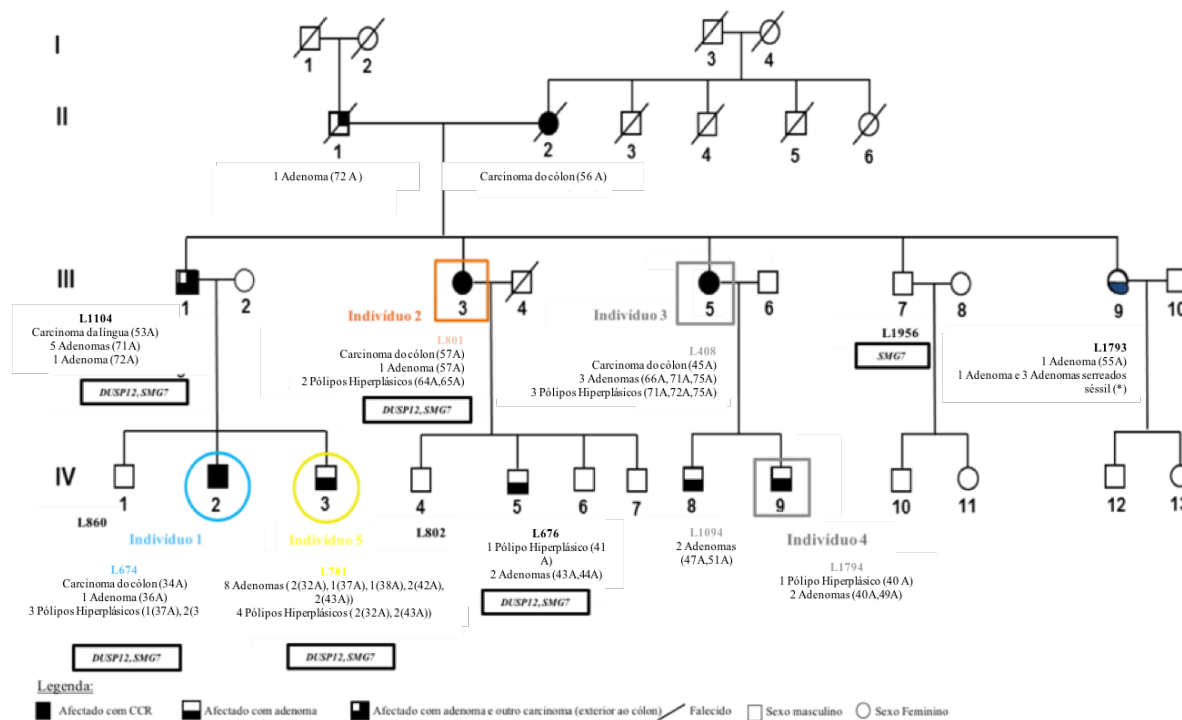


Figura 4.7. Os indivíduos assinalados a cores foram selecionados para análise bioinformática, no que diz respeito à análise 5. Apenas para os indivíduos com número de identificação LXXX ou LXXXX existe material biológico disponível. Para os indivíduos afectados foi feita uma caracterização fenotípica e respectiva idade de diagnóstico com a excepção de * para o qual não se dispunha de informação disponível. Todos os adenomas identificados correspondem a adenomas tubulares com displasia de baixo grau, com a excepção no indivíduo L1104 que apresenta uma adenoma tubular com displasia de alto grau aos 71A.

Para além destas variantes seleccionadas, o estudo de segregação ainda contemplou a presença de mais 3 variantes resultantes de uma análise bioinformática preliminar anterior à análise bioinformática subjacente a este estudo, em que incluía as variantes somente partilhadas pelos indivíduos 1,2,3 e 4.

Da referida análise bioinformática preliminar a este estudo tinham sido identificadas três variantes possivelmente patogénicas que também segregavam na família, nomeadamente nos genes *RNF207* (p.Arg539Cys), *ZBTB48* (p.Ser675Ala) e *TASIR1* (p.Leu274Phe) (figura 4.8). Estas variantes foram identificadas na grande maioria dos indivíduos afectados (excepto L1793) e no indivíduo saudável L1956.

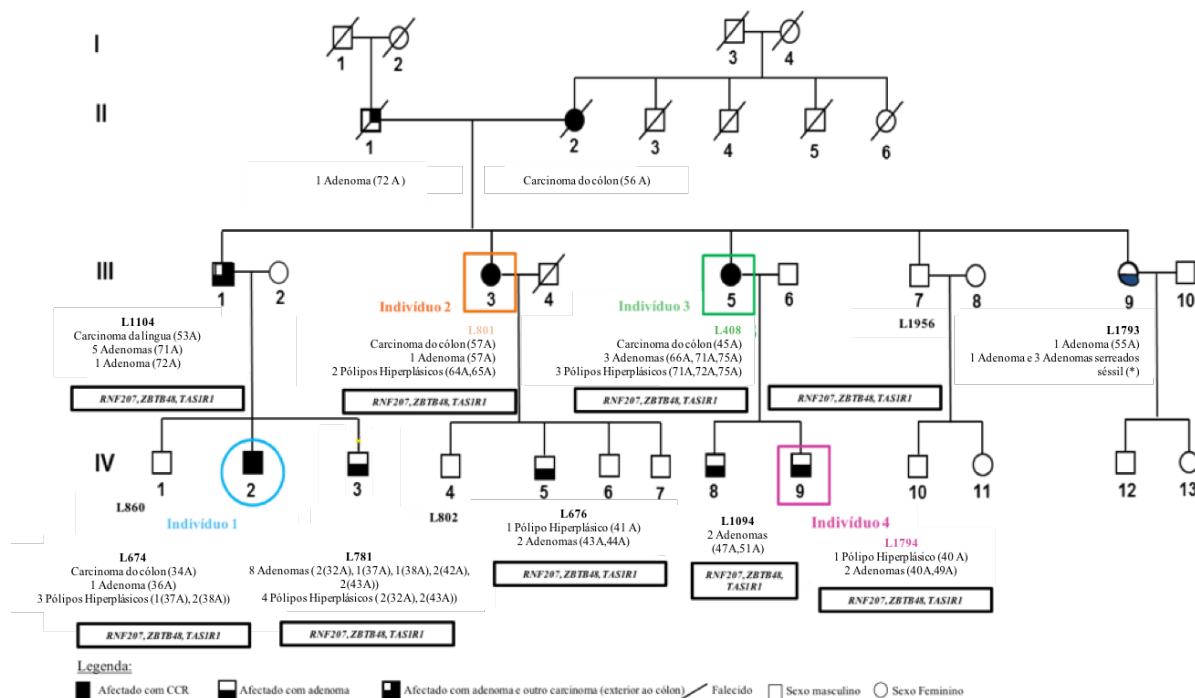


Figura 4.8. Os indivíduos assinalados a cores foram seleccionados para análise bioinformática, no que diz respeito à análise bioinformática previamente realizada ao estudo. Apenas para os indivíduos com número de identificação LXXX ou LXXXX existe material biológico disponível. Para os indivíduos afectados foi feita uma caracterização fenotípica e respectiva idade de diagnóstico com a excepção de * para o qual não se dispunha de informação disponível. Todos os adenomas identificados correspondem a adenomas tubulares com displasia de baixo grau, com a excepção no indivíduo L1104 que apresenta um adenoma tubular com displasia de alto grau aos 71A.

A localização cromossómica dos vários genes onde foram identificadas as variantes referidas encontra-se representada na figura 4.9.

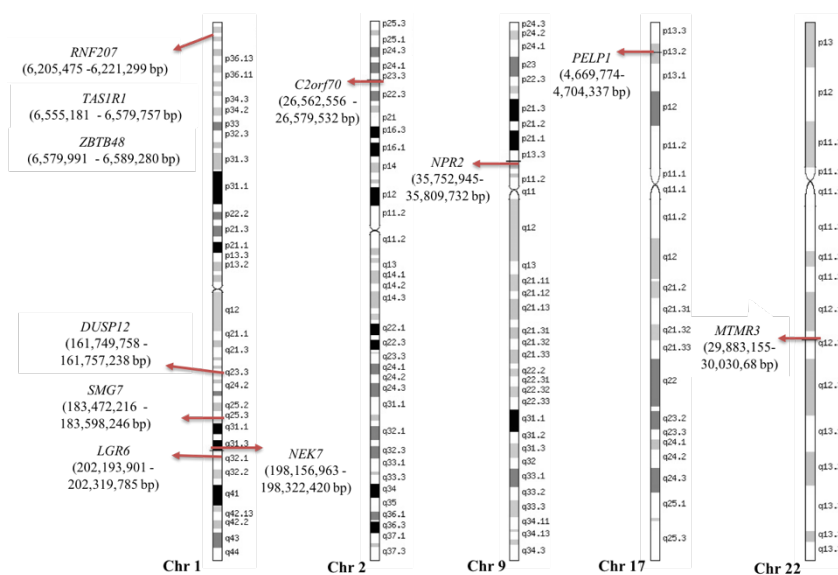


Figura 4.9. Representação cromossómica dos vários genes onde foram identificadas as variantes. (Adaptado de: Genecards (<http://www.genecards.org/>))

4.3.3. Análise somática da variante p.Arg645Trp identificada no gene *MTMR3*

Sendo a variante p.Arg645Trp identificada no gene *MTMR3*, a que mostrou uma maior segregação colocou-se a hipótese de no caso de se confirmar como associada à susceptibilidade para o FCCTX na família e funcionar como gene supressor neste modelo de tumorigénese, o gene poder apresentar um segundo evento, por exemplo perda de heterozigotia. A registar-se perda de heterozigotia, a análise da variante no tumor por sequenciação deveria revelar a variante em homozigotia devido a deleção do alelo normal. Neste caso, a análise somática realizada em amostras de tecido incluído em parafina para 4 indivíduos afectados da família FCCTX em estudo confirmou a presença da variante p.Arg645Trp em heterozigotia. Independentemente da lesão no epitélio do cólon (pólipo, adenoma ou carcinoma) a variante foi sempre confirmada em heterozigotia. Para a amostra do indivíduo não pertencente à família, a variante não foi observada (tabela 4.12).

Tabela 4.12. Resultados obtidos para a análise somática da variante p.Arg645Trp do gene *MTMR3*.

Indivíduo da família L56	Amostra de parafina correspondente	Caracterização do tecido incluído em parafina	Identificação da variante p.Arg645Trp
L674	Cas 995	Carcinoma	Heterozigotia
L676	L676P	Pólipo	Heterozigotia
L781	Cas 987	Adenoma Tubular	Heterozigotia
L801	Cas 3575	Carcinoma	Heterozigotia
Não pertencente à família	Cas 595	Carcinoma	Ausente

4.3.4 Análise de expressão do gene *MTMR3* e *RAC1* no tumor de um indivíduo portador da mutação germinal neste gene

Tendo em conta a interacção descrita entre as proteínas MTMR3 e RAC1, mediada pela interacção do PtdIns5P um produto de desfosforilação da proteína MTMR3 (Oppelt et al., 2014) ; o facto destes dois genes e a isoforma do *RAC1* (*RAC1b*) já terem sido descritos como intervenientes no desenvolvimento tumoral no CCR, colocou-se a hipótese de a expressão destes genes poder estar alterada em tumores de indivíduos portadores de mutação germinal no gene *MTMR3*. Os resultados da análise de expressão sugerem uma expressão dos genes *MTMR3*, *RAC1* e *RAC1b* no tumor FCCTX em comparação com as mucosas normais B e C. Contudo, também se observou uma expressão elevada destes genes em tumores não FCCTX (A193C e CAS638C). A linha celular humana de cancro do

cólon e recto SW480 revelou uma sub-expressão em comparação com as restantes amostras (figura 4.10).

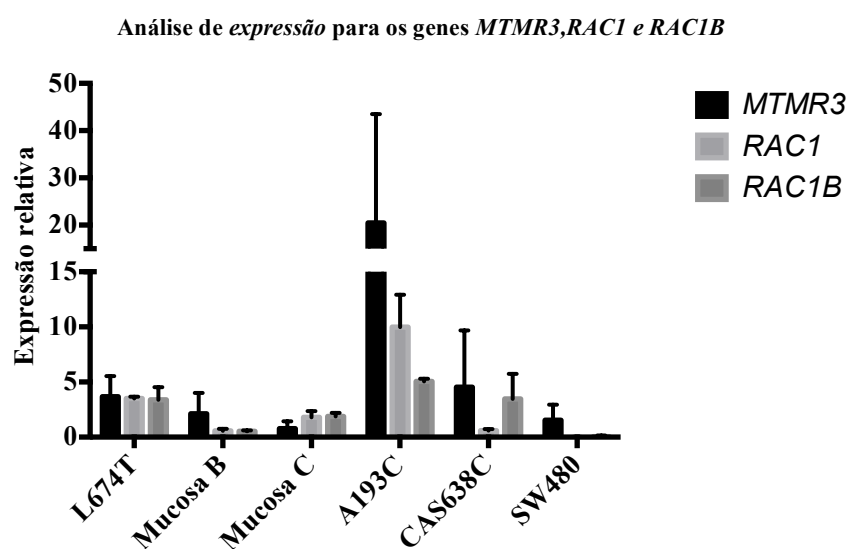


Figura 4.10. Análise de expressão para os genes *MTMR3*, *RAC1* e *RAC1b* num tumor de um indivíduo FCCTX portador de mutação no gene *MTMR3*, em dois tumores não associados ao FCCTX , numa linha celular e em mucosa cólica normal.

5. Discussão

5.1. Análise do envolvimento do gene *FBXO11* na susceptibilidade para FCCTX

A análise mutacional do gene *FBXO11* realizada na linha germinal de indivíduos FCCTX permitiu analisar um possível envolvimento deste gene nesta síndrome. Desta análise, foram apenas identificadas variantes sem potencial carácter patogénico. Mesmo no caso da variante *missense* (p.Thr126Ser) detectada esta é classificada como benigna pela análise *in silico*, pelo que juntamente com a sua elevada frequência, esta variante não pode ser considerada como a causadora do fenótipo de FCCTX. A variante sinónima também pode ser excluída como causadora de patogenicidade visto que a análise *in silico* a classifica como polimorfismo e não revela potenciais alterações de locais de *splicing*.

Apenas a variante c.2555+53T>C (intrão 21) poderia justificar a susceptibilidade deste gene para FCCTX devido à baixa frequência mundial associada às mesmas. Do grupo de indivíduos em estudo com tumores sem expressão de *MSH6*, só para o L1152 foi identificada a alteração c.2555+53T>C. No entanto, a sua localização não sugere que possam ser patogénicas.

Toda os estudos efectuados, desde a análise mutacional à análise de *copy number* e de expressão efectuados para o gene *FBXO11* não identificaram nenhuma alteração neste gene que o formulasse como envolvido na susceptibilidade para FCCTX. Não se pode contudo excluir o seu envolvimento na tumorigénese colorectal a nível somático tendo em conta a sua interacção com o gene *MSH6*. Estudos adicionais serão necessários para avaliar esta hipótese.

5.2. Caracterização a nível germinal, das famílias FCCTX (subgrupos TSG⁺ e TSG⁻) através da análise de um painel multigénico por *Next Generation Sequencing*

A análise de mutações germinais nos indivíduos de famílias FCCTX, realizada através do painel multigénico (*Trusight Cancer*) permitiu identificar 8 variantes apenas em famílias TSG⁺, em genes associados à via de reparação HR que é essencial para manter a integridade do genoma após erros no DNA do tipo DSB (*double-strand break*).

Uma das 8 variantes identificadas (p.Lys332Arg) localiza-se no gene *RECQL4*. A proteína codificada por este gene, intervém nesta via de reparação por promover o recrutamento das proteínas CtiP, BLM (*Bloom Syndrome RecQ Like Helicase*), DNA2 (*DNA Replication Helicase/Nuclease 2*) e EXO1 (*Exonuclease 1*) intervenientes na remodelação da sequência (Breaks et al., 2016). Desta forma esta mutação pode afectar este processo de reparação e contribuir para aumentar o risco de CCR.

Um dos processos mais frequentes após a indução de DSB é a remodelação da cromatina através da ubiquitinação das proteínas H2AX (*Histone family member X*) e H2AB (*Histone cluster 2 H2A family member B*), de maneira a deixar acessível os sítios danificados para reparação (Chenevix-trench et al.,

2014). A ubiquitinação da proteína H2AB é mediada por um complexo formado pelas proteínas RNF20 (*Ring finger protein 20*) e RNF40 (*Ring finger protein 40*) (Zhu et al., 2005). Num estudo conduzido por Hahn et al., 2012 revelou que o CDC73 é essencial para a estabilidade do complexo RNF20-RNF40 e que mutações patogénicas neste gene podem conduzir à perda da actividade deste complexo. Uma variante identificada no gene *CDC73* (p.Arg1445Trp) foi detectada no indivíduo L48 pertencente à família 24. Neste indivíduo já se tinha observado uma deleção dos exões 5-8 do gene *SMAD4*. A perda da função do gene *SMAD4* está associado ao desenvolvimento tumoral e a um pior prognóstico em CCR (Yan et al., 2016). O fenótipo apresentado por este doente é muito agressivo, com o desenvolvimento de tumores extra-cólicos para além do CCR. Deste modo, não é de excluir que as 2 alterações nestes genes possam contribuir para o fenótipo.

O indivíduo L108 apresenta duas mutações, uma no gene *FANCA* (c.10007-3C>A) detectada nesta análise do painel e outra no gene *POLD1* (p.Pro300Leu), identificada anteriormente num outro estudo. A variante descrita para o gene *FANCA* está associada a uma perda de local de aceitador de *splicing* (*Mutation Taster*), o que poderá implicar uma proteína aberrante incapaz de corrigir os erros de DSB. O gene *POLD1* também faz parte de um sistema de reparação de erros inseridos durante a replicação do DNA (Palles et al., 2013), pelo que qualquer uma ou mesmo a combinação das duas variantes poderá contribuir para o FCCTX nesta família.

Também no caso do indivíduo L84 já tinha sido identificada previamente uma mutação do tipo *missense* no gene *APC* (p.Glu1317Gln) (Francisco et al, 2011), tendo a análise recorrendo ao painel multigénico vindo a identificar uma mutação no gene *WRN*. Este gene codifica para uma proteína com função de helicase a qual intervém igualmente na reparação de DSB por recombinação homóloga (Chen et al., 2003). Desta forma a mutação no gene *WRN* pode, juntamente com a mutação no gene *APC*, aumentar o risco de adenomas e contribuir para o FCCTX nesta família.

O indivíduo L984 revelou a presença de uma mutação (p.Arg370Pro) no gene *WT1*. Este gene já foi igualmente associado à HR através da interacção de uma das suas isoformas (WTC) com proteínas intervenientes nesta via tais como as proteínas XRCC2 (*X-ray repair cross complementing 2*), Rad51D (*RAD51 paralog D*) e Rad54L (*RAD54-like (S.cerevisiae)*) (Oji et al., 2015).

As mutações p.Gly439Val e p.Ser1434Leu dos genes *FANCE* e *SLX4*, respectivamente, foram identificadas para o mesmo indivíduo, L1038, pertencente à família 173. Sendo que estes genes já foram associados à HR, as duas mutações em conjunto intensificam os erros de reparação associados a esta via (Renato et al., 2016) (Michl et al., 2016).

Para além disto, foi ainda identificada uma mutação no gene *XPC* (p.Phe287Cys), também associado a reparação de erros no DNA, para o indivíduo L1200 (Laczmanska et al., 2007).

Para o indivíduo L451 não foi detectada nenhuma mutação através da análise do painel, o que implica que a sua susceptibilidade para FCCTX poderá não estar associada a genes que participam nas vias de reparação de erros no DNA. Neste mesmo doente já foi observada uma mutação no gene

BMPRIA, mais concretamente no exão 10 do mesmo (p.Leu157Pro), sendo esta a razão mais plausível de justificar o fenótipo observado no indivíduo.

Os indivíduos que apresentaram mutações em genes de reparação associados à HR apresentaram tumores classificados como TSG⁺, ou seja caracterizados por apresentarem metilação nos genes MMR e *MGMT* e mutações somáticas nos genes *APC* e *KRAS*. Se nestes indivíduos os efeitos genotóxicos nas células epiteliais do cólon não forem reparados pelos sistemas de reparação MMR e *MGMT* conduzem à inserção de quebras em dupla cadeia no DNA (figura 5.1). O defeito germinal no sistema de reparação HR pode provocar assim um aumento da probabilidade de alterações cromossómicas e uma instabilidade genómica, contribuindo assim para o desenvolvimento tumoral e associar-se à susceptibilidade aumentada nestas famílias FCCTX.

Do sub-grupo de doentes de famílias HNPCC ou que preenchem os CB, a maioria das mutações identificadas também foram identificadas em genes associados à HR. Dos genes em comum entre este grupo e os indivíduos índice de famílias FCCTX estão os genes *SLX4*, *XPC* e *WRN*. Estes resultados vieram corroborar a hipótese de estes sub-grupos de doentes poderem ser potenciais FCCTX, visto que se identificaram mutações comuns e que são casos HNPCC (muitos deles terão tumores MSS e virão a ser classificados como FCCTX) e CB. É de notar que um número substancial dos indivíduos CB com mutação germinal nestes genes de reparação preenchem os CB4 ou CB5, que correspondem a uma história familiar de CCR, no entanto incompleta e por isso não preenchem os critérios de Amesterdão.

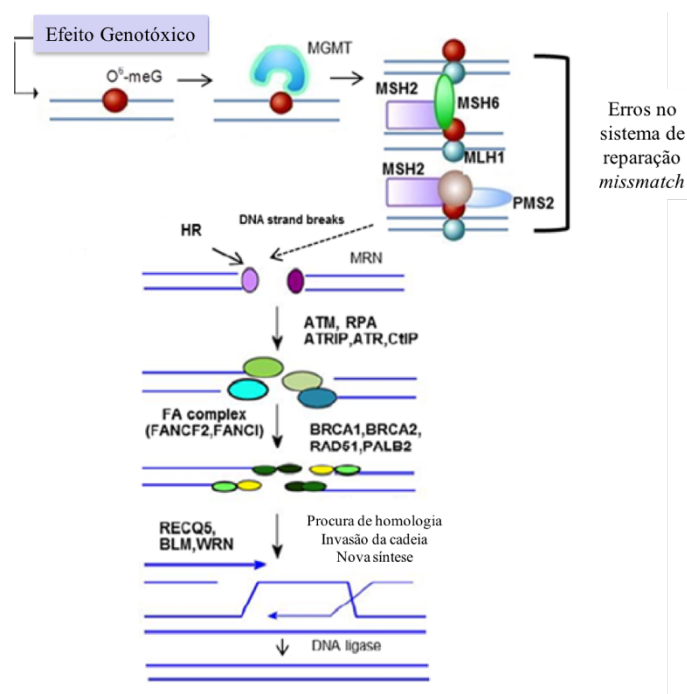


Figura 5.1. Mecanismo sugerido pelo presente estudo, envolvido na reparação de erros no DNA causados por stress genotóxico em FCCTX. (Adaptado de : (Liu e Gerson, 2006)(Peng e Lin, 2011))

Em suma, mutações germinais nos genes de reparação do DNA por HR podem contribuir para o aumento do risco de tumores do cólon e recto num subgrupo de famílias FCCTX, principalmente com tumores TSG⁺. Estes resultados vêm confirmar uma divisão do FCCTX em várias entidades, nas quais estarão patentes diferentes causas moleculares.

O facto de em algumas famílias FCCTX se terem identificado duas variantes provavelmente patogénicas, ambas bastante sugestivas de se associarem à doença, sugere como foi já sugerido que a combinação de mutações de baixo e moderado risco possam contribuir para a agregação familiar (Dominguez-Valentin *et al.*, 2015). Este facto, apoia a hipótese que em muitos casos esta síndrome não seja monogénica, mas antes poligénica, resultando da interacção de vários genes de baixa penetrância (Stoffel e Kastrinos, 2014).

5.3 Identificação de novos genes candidatos para a susceptibilidade para o FCCTX por sequenciação do exoma

A análise de exoma efectuada para a família L56 permitiu identificar variantes que segregassem na família de modo a justificar a susceptibilidade para o FCCTX. Dos resultados da análise de segregação, identificou-se a variante p.Arg645Trp, do gene *MTMR3*, como possivelmente patogénica, revelando segregação com a doença na família. A análise de segregação, tendo em conta os diferentes modelos de análise, revelou ainda 7 variantes que segregavam parcialmente: *NEK7* (p.Arg356Gly), *C2orf70* (p.Pro78Ala), *NPR2* (p.Val882Ile), *LGR6* (p.Arg623Gln), *PELP1* (p.Ser298=), *DUSP12* (p.Leu2_Leu3ins) e *SMG7* (p.Pro657His).

A variante identificada para o gene *MTMR3* (p.Arg645Trp), foi a única que resultou da análise 1 que engloba as variantes partilhadas pelos 5 indivíduos afectados para os quais foi feita análise de WES. Esta variante está descrita com o SNP rs138823197, classificada como *deleterious/probably damaging* nos softwares *SIFT/Polyphen* e apresenta uma frequência mundial de aproximadamente 0,02%. Na família L56, a alteração foi identificada nos indivíduos afectados quer com carcinoma quer com adenomas com a excepção do indivíduo L1094 que até ao momento é o que apresenta um fenótipo menos severo (presença de apenas dois adenomas) e não se pode excluir se não poderão ter sido desenvolvidos no contexto esporádico.

Na base de dados *COSMIC* estão descritas 184 variantes para este gene em que 17 estão presentes em CCR (v78 <http://cancer.sanger.ac.uk/cosmic>). Para este gene também já foram descritas mais 78 variantes na base de dados *Cancer Cell Metabolism Gene* (<https://bioinfo.uth.edu/ccmGDB/index.html>), em que 12 são em CCR.

A proteína *MTMR3* é responsável pela desfosforilação do fosfatidilinositol PtdIns3P e do PtdIns(3,5)P₂ para PtdIns e PtdIns5P, respectivamente (figura 5.2-A) (Taylor et al., 2000)(Blondeau et al., 2000). Para executar a sua função, a proteína necessita de 4 domínios funcionais: o domínio PH-

GRAM para ligação ao substrato, o domínio de fosfatase PTP para a sua acção enzimática, o domínio *Coiled-coil* para uma heterodimerização/homodimerização da proteína que define o seu estado activo e inactivo e o domínio Zinc finger que permite uma interacção entre proteínas (figura 5.2-B) (Hnia et al., 2012).

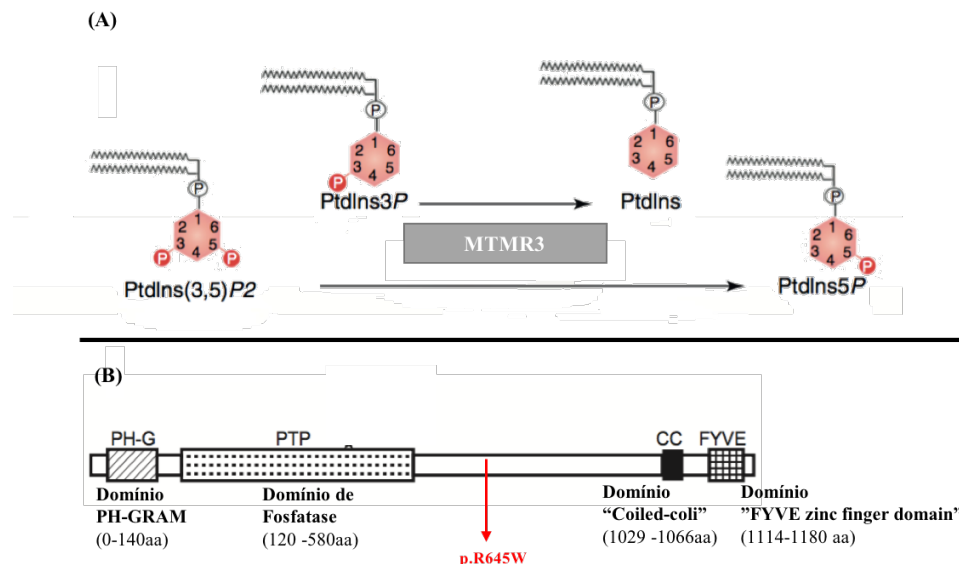


Figura 5.2. (A) Representação da desfosforilação do fosfatidilinositol PtdIns3P e do PtdIns(3,5)P2 pela MTMR3; (B) Representação dos domínios proteicos da MTMR3 e a localização da variante identificada no contexto da análise de exoma. (adaptado de : (Lorenzo et al., 2005) (Hnia et al., 2012))

Apesar da localização da variante não ser coincidente com nenhum domínio proteico, a variante está presente num local rico de sítios de fosforilação (figura 5.3). A actividade da proteína poderá ser dependente de um processo de ubiquitinação como a proteína da sua família MTMR4 (Lorenzo et al., 2005). Para esta última foi identificado o mecanismo pelo qual esta é degradada para proteossoma mediado pela proteína Nedd4 que tem capacidade de ligação à ubiquitina E3 e ligação à MTMR4 no seu domínio PY(Pro-TYR). A actividade da proteína MTMR3 pode estar comprometida por um processo semelhante visto que a alteração do aminoácido pode conduzir a uma alteração de estrutura proteica deixando expostas ou escondendo sequências reconhecidas por proteínas como a Nedd4 e assim afectar a sua marcação para degradação (Plant et al., 2009).

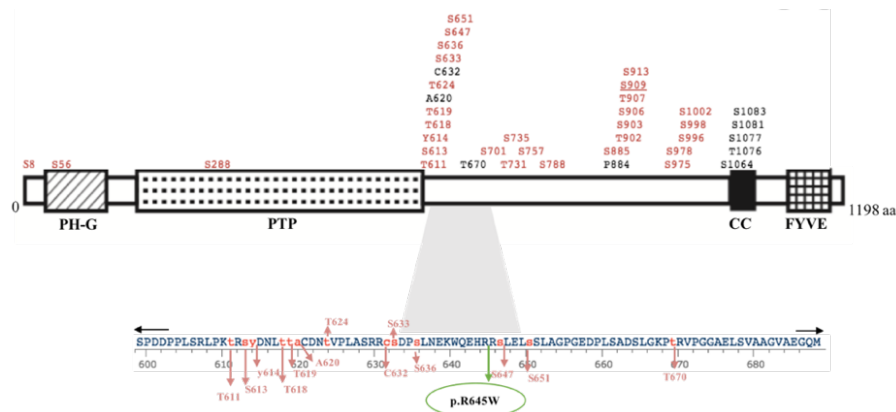


Figura 5.3. Representação dos domínios proteicos com os vários sítios de fosforilação (vermelho) da proteína e a variante identificada pela análise de exoma. (Adaptado de : (Lorenzo et al., 2005) (PhosphoSitePlus: <http://www.phosphosite.org/homeAction.action>))

As funções já definidas para esta proteína envolvem a sua participação no processo de autofagia e na via de sinalização mTOR. A autofagia é um processo de degradação de moléculas regulada pelos níveis PtdIns3P, um dos alvos da proteína MTMR3, consoante o estado nutricional das células. Assim, níveis baixos da proteína MTMR3 permitem aumentar os níveis de PtdIns3P permitindo que esta recrute a WIP1-1 para iniciação da autofagia. Por outro lado, a proteína tem a capacidade de fosforilar a proteína mTORC1 tornando-a inactiva e assim induzir a autofagia (Taguchi-Atarashi et al., 2010) (Hao et al., 2016).

No trabalho realizado por Oppelt et al., 2014 demonstraram que durante a migração celular, a produção de PtdIns5P levado a cabo pela proteína MTMR3 leva ao recrutamento de proteínas efectoras (possivelmente GEFS-*guanine-nucleotide-exchange factors*) que recrutam a proteína RAC1. O RAC1 contribui para a proliferação celular, invasão e para o processo metastático (Espina et al., 2008). A sua isoforma alternativa RAC1B também já foi identificada em tumores CCR de forma *up-regulated* em comparação com a mucosa normal. Esta isoforma pode também contribuir para a progressão tumoral de modo a ultrapassar alguns efeitos antagonistas gerados pelos efectores da RAC1 (Matos e Jordan, 2008). Na análise de expressão realizada no presente estudo, revelou-se que ambos os três genes se encontravam sobre-expressos na amostra tumoral FCCTX (L674T) em comparação com mucosas normais B e C de indivíduos não afectados. No entanto, serão necessários mais estudos incluindo a análise de expressão noutros tumores da família e das respectivas mucosas normais. Os resultados tanto da análise de expressão como da análise mutacional indicam uma possibilidade deste gene ser um candidato para a susceptibilidade para FCCTX. Porém, é necessário detalhar a sua função na participação da progressão tumoral desta síndrome. Para tal, já foram planeados testes funcionais *in vitro* de modo a avaliar o efeito patogénico desta variante em processos celulares como a apoptose, viabilidade e migração celular.

Da análise 3 que inclui as variantes partilhadas pelos indivíduos 1 (L674), 3 (L408), 4 (L1794) e 5 (L781) com a exclusão do indivíduo 2 (L801), embora pouco provável, como possível fenocópia, a variante identificada no gene *NEK7* revelou segregação parcial na família e identificada em indivíduos

com aparente predisposição para o desenvolvimento de adenomas. Esta variante já se encontra descrita na base de dados NCBI com o SNP rs558333332, no entanto pelos scores *SIFT/Polyphen* é considerada como *tolerated/benign*. Para além disto, esta variante apresenta uma frequência mundial de 0,4% e uma frequência europeia perto dos 0,6%. Quanto à função que desempenha, esta proteína intervém na progressão do ciclo celular, através da formação do fuso mitótico (figura 5.4). Uma possível ligação entre a NEK7 e o desenvolvimento tumoral poderia passar pela via CIN através de uma possível desregulação da proteína conduzir a uma deficiente segregação cromossómica levando a alterações numéricas na célula (Regan e Fry, 2009). O seu envolvimento no aumento do risco de CCR ou adenomas é por isso incerto, necessitando de estudos adicionais.

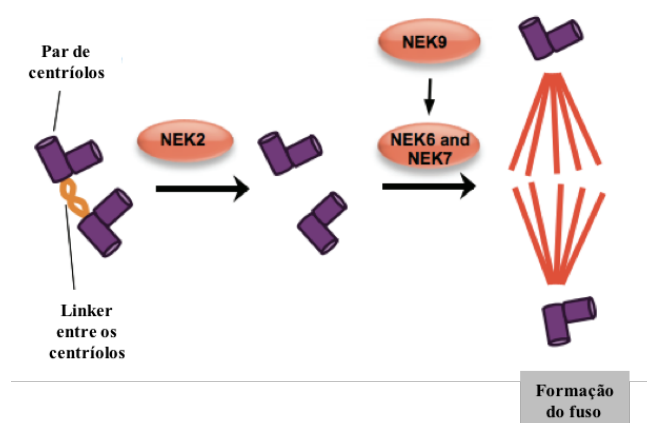


Figura 5.4. representação esquemática do modelo de função da proteína NEK7 no ciclo celular. Após a fosforilação das proteínas NEK6 e NEK7 por parte da NEK9, estas fosforilam vários componentes essenciais para a formação do fuso mitótico. (Adptado de:(Fry et al., 2012))

Face ao modelo da análise 4, as variantes seleccionadas nesta análise eram partilhadas apenas pelos indivíduos 1 (L674), 3 (L1408) e 5 (L781), considerando como casos esporádicos os indivíduos 2 (L801) e 4 (L1794). Desta análise foram identificadas 4 variantes a segregar parcialmente com a doença na família nos genes *C2orf70*, *NPR2*, *LGR6* e *PELP1*.

A variante do tipo *missense* p.Pro78Ala identificada no gene *C2orf70* encontra-se descrita na base de dados NCBI com o SNP rs112873448 e apresenta uma frequência mundial de 0,7%. Os softwares de *SIFT/Polphen* prevêm a alteração deste aminoácido como *deleterious/probably* e a sua frequência europeia equivale a 1,8%. Tendo em conta toda esta informação *in silico*, nomeadamente a sua elevada frequência europeia, e que a função desta proteína continua uma incógnita, esta variante não deverá estar associada ao fenótipo FCCTX na família.

Para o gene *NPR2* foi identificada a variante p.Val882Ile com segregação parcial na família. Este gene é responsável por codificar a enzima transmembranar GB-C (*Guanylate Cyclase B*) que catalisa a conversão do GTP (*guanosine triphosphate*) em cGMP (*cyclic GMP*) (figura 5.5).

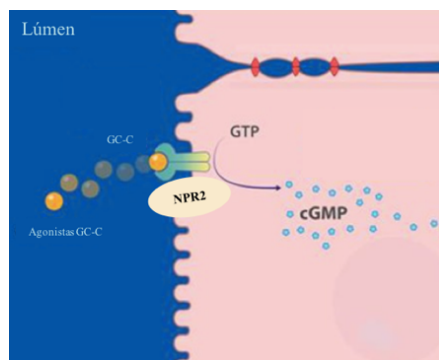


Figura 5.5. Possível mecanismo subjacente à conversão de GTP em cGMP pela proteína NPR2. (Adaptado de : Hannig et al., 2014)

A proteína NPR2 é constituída por 4 domínios essenciais: o domínio CYC (*Cyclase catalytic domain*), o domínio ECD (*Extracellular-binding domain*), o domínio TM (*Transmembrane segment*) e o domínio KHD (*Protein Kinase homology domain*) (figura 5.6) (Tamura e Garbers, 2003).

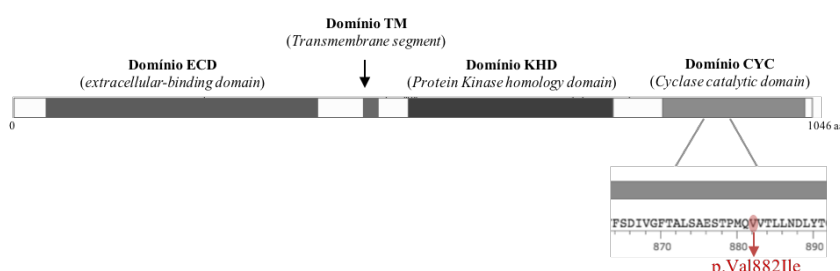


Figura 5.6. Representação dos domínios proteicos do GC-B e identificação da variante p.Val882Ile na respectiva região do domínio CYC da proteína NPR2. (Adaptado de : (Tamura e Garbers, 2003) (PhosphoSitePlus: <http://www.phosphosite.org/homeAction.action>))

A alteração identificada na análise localiza-se num dos 4 domínios essenciais da proteína, o domínio (CYC), na posição 882 provocando a alteração de uma valina por uma isoleucina. Estudos recentes identificaram uma alteração no aminoácido adjacente (p.Val883Met) como causadora de uma modificação a nível da conformação da proteína, levando a uma produção constitutiva de cGMP. Os níveis de cGMP são essenciais para uma homeostasia normal no epitélio do cólon, equilibrando os processos de diferenciação e proliferação das células na cripta (Miura et al., 2012). Por estas razões, se a variante p.Val882Ile também for capaz de modificar a conformação da proteína, implica níveis de expressão aberrantes do cGMP, o que pode induzir alterações nos processos celulares referidos, contribuindo para o desenvolvimento de alterações no epitélio do cólon. Para além disto, a variante já

se encontra descrita com o SNP rs55700371, apresenta uma baixa frequência mundial e europeia e é considerada como *deleterious/possibly damaging* pelos softwares *SIFT/Polyphen*. Esta hipótese também é reforçada pelo facto de os indivíduos que apresentam esta variante no gene *NPR2*, à excepção de um, todos terem desenvolvido carcinomas. O indivíduo que ainda não desenvolveu carcinoma, apresenta um número elevado de pólipos em idade muito jovem e possivelmente não desenvolveu carcinoma como o irmão porque entretanto começou a ser vigiado.

O gene *LGR6* é responsável por codificar um receptor acoplado a proteína G. Estudos baseados na caracterização molecular desta proteína, comprovaram a afinidade deste receptor para a família das proteínas RSPO (*R-spondin*), semelhante aos outros receptores da mesma família, LGR4 (*Leucine rich repeat containing G protein-coupled receptor 4*) e LGR5 (*Leucine rich repeat containing G protein-coupled receptor 5*). A interacção do receptor LGR6 com uma proteína RSPO é responsável por activar a via de sinalização *WNT* que intervém na proliferação e migração celular no cólon (figura 5.7). Adicionalmente à variante identificada neste estudo, outras 3 mutações neste gene em 37 casos de CCR também foram identificadas num contexto de sequenciação de exoma. Para além disto, a região promotora do gene já foi encontrada hipermetilada num estudo envolvendo casos de CCR (Sjöblom et al., 2006)(Mokarram et al., 2009). A variante p.Arg623Gln identificada para este gene está localizada num domínio 7TM (domínio transmembranar) necessário na via de sinalização do receptor. Apesar dos softwares *SIFT/Polyphen* preverem esta alteração como *tolerated/benign* (perto do limiar de *score*) a sua frequência mundial (0,002%) e toda a informação funcional prevista para a proteína LGR6 fortalecem uma possível patogenicidade associada a esta variante. Para além disto, esta variante está presente em indivíduos com um número considerável de adenomas, o que poderá reforçar a sua associação a um risco aumentado de adenomas e poder associar-se ao fenótipo apresentado por esta família.

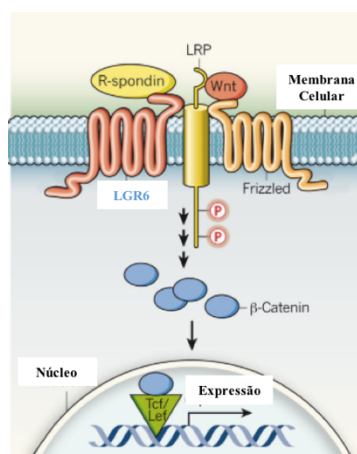


Figura 5.7. Representação da possível interacção do LGR6 com a RSPO e a sua participação na via *WNT*. (Adaptado de Birchmeier, 2011)

Pela análise 4 ainda foi identificada a variante p.Ser298= no gene *PELPI*. Esta variante também já se encontra descrita na base NCBI com o SNP rs68343383 e com uma frequência mundial e europeia associada de 0,04% e 0,03%. Visto que esta proteína é do tipo sinónima não vai implicar alteração de aminoácido mas pode ter implicações a nível de *splicing*. Pelo software *Mutation Taster* a alteração da citosina (C) por uma tirosina (T) não demonstra alterações relevantes a nível de *splicing*. Para este gene já foram descritas duas isoformas: uma isoforma longa de 3,8Kb e outra mais curta de 3,4 Kb, em que esta última é expressa em várias células tumorais. O *PELPI* já foi definido como proto-oncogene e a sua expressão já foi determinada em tumores de cólon. Quando este exerce a sua função de oncogene ele influencia vários processos como o ciclo celular, modificações na cromatina, metilação de histonas e resposta a danificações na sequência de DNA (Girard et al.,2014). No entanto, a previsão *in silico* não favorece a sua patogenicidade, assim como a sua presença no indivíduo L1956, o qual não é afectado.

A análise 5 englobou as variantes partilhadas pelos indivíduos 1 (L674), 2 (L801) e 5 (L781) especulando os indivíduos 3 (L408) e 4 (L1794) como esporádicos. Desta análise derivaram duas variantes para dois genes: *DUSP12* e *SMG7*.

A variante p.Leu2_Leu3ins foi identificada no gene *DUSP12* que codifica uma fosfatase responsável por desfosforilar vários resíduos de tirosina, serina ou treonina (Patterson et al.,2009). Apesar desta proteína ainda não ter função completamente conhecida já foi descrita uma possível interacção desta com a proteína HSP70 (*heat shock protein*) que actua em vias dependentes de caspases para inibir apoptose (figura 5.8). Esta proteína ainda pode ter a possibilidade de promover o desenvolvimento tumoral através da interacção com dois oncogenes *ITGA1* (*Integrin Subunit Alpha 1*) e *c-MET* (*MET Proto-Oncogene, Receptor Tyrosine Kinase*) que participam na invasão tumoral e migração celular, respectivamente (Cain e Beeser, 2013). A *frameshift* identificada ainda não está descrita na base NCBI e ainda não foi reportada nenhuma frequência associada à mesma, o que poderá reforçar uma eventual patogenicidade.

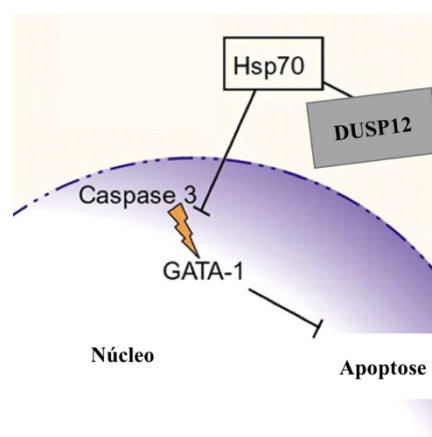


Figura 5.8. Representação da interacção das proteínas DUSP12 e HSP70 e o seu possível envolvimento na inibição da apoptose. (Adaptado de Weiss et al., 2009)

Pelo contrário, a variante p.Pro657His, identificada no gene *SMG7*, já se encontra descrita com o SNP rs34426362 com uma frequência de 0,2% e 0,6% a nível mundial e europeia, respectivamente. Para além disto, esta variante é considerada pelos *softwares SIFT/Polyphen* como *deleterious/probably*. Num estudo conduzido por Luo et al., 2016 foi revelada uma função relevante da proteína SMG7 na estabilização do p53 em resposta aos erros na sequência de DNA. A proteína SMG7 através da fosforilação dos resíduos de serina na proteína Mdm2 leva a que esta não se ligue ao p53 para inibir a sua actividade e poder reparar os erros de DNA (Luo et al., 2016). A segregação destas duas variantes revelou-se em conjunto pois estavam presentes nos mesmos indivíduos afectados (L1104, L801, L674, L781, L676). Este haplotipo partilhado na família deve-se ao facto de estes genes estarem localizados em regiões adjacentes no cromossoma 1 e poderia estar associado a um agravamento de fenótipo visto que os indivíduos detectados com o mesmo são os que apresentam um fenótipo mais marcado. No entanto, a variante identificada no gene *SMG7* também se encontra presente no indivíduo L1956 que não é afectado, pelo que talvez favoreça a variante detectada no gene *DUSP12* como mais provavelmente associada ao fenótipo nesta família.

Da análise bioinformática anteriormente ao estudo foram identificadas três variantes presentes nos genes *RNF207* (p.Arg539Cys), *ZBTB48* (p.Ser675Ala) e *TAS1R1* (p.Leu274Phe). Estas variantes foram descritas para os mesmos indivíduos afectados e num único indivíduo saudável (L1956) devido às localizações vizinhas dos três genes no cromossoma 1 (figura 4.9). As variantes p.Arg539Cys e p.Ser675Ala já estão descritas na base NCBI com o SNP rs55823245 e rs2229330, respectivamente, apresentam uma frequência mundial (0,9% e 3,5%) e europeia (2,6% e 7,2%) elevada e estão caracterizadas como *deleterious/possibly damaging* e *deleterious/benign* nos *softwares SIFT/Polyphen*. A previsão *in silico* destas duas variantes indicam uma fraca contribuição das mesmas na susceptibilidade para FCCTX na família. Por outro lado, a variante identificada no gene *TAS1R1* (p.Leu274Phe) ainda não apresenta nenhum SNP atribuído na base NCBI. A proteína codificada por este gene, um receptor acoplado à proteína G, está associada principalmente à percepção do paladar humano e na detecção de aminoácidos ao longo do tracto gastrointestinal. Estudos recentes demonstraram uma possível ligação entre este receptor e a proteína MTORC1, e assim à semelhança da proteína MTMR3, pode estar envolvido no processo de autofagia através da via de sinalização mTOR (figura 5.9) (Choi et al., 2016) (Wauson et al., 2012).

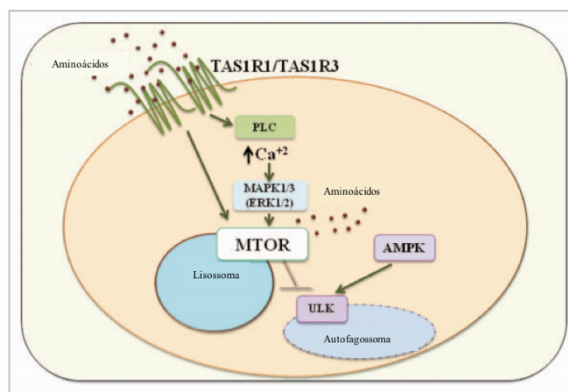


Figura 5.9. Acção dos receptores TAS1R1/TAS1R3 na presença de aminoácidos e a sua consequente activação da proteína MTORC1 de forma a inibir a autofagia. (Adaptado de Wauson et al., 2013)

Em suma, após a selecção criteriosa de variantes potencialmente patogénicas e reunindo informações sobre a sua segregação na família (face ao modelo onde se inserem), funções relevantes em desenvolvimento tumoral e patogenicidade prevista *in silico*, formularam-se 3 hipóteses que poderão estar subjacentes à susceptibilidade de FCCTX nesta família. A primeira hipótese baseia-se na segregação da variante identificada no gene *MTMR3* e no facto deste ter-se revelado um gene bastante promissor e com grande impacto para o desenvolvimento de CCR, implicando que pode estar por detrás do fenótipo associado a esta família. A segunda hipótese é que juntamente com a mutação no gene *MTMR3*, mutações nos genes *LGR6*, *NEK7* ou *TAS1R1* poderão revelar um aumento de risco para desenvolvimento de adenomas. Por último, as variantes observadas nos genes como o *DUSP12* e o *NPR2* poderão intervir como potenciais modificadores de fenótipo.

6. Conclusão

O cancro do cólon e recto familiar (FCCTX) não apresenta ainda causas moleculares conhecidas que expliquem a susceptibilidade para o desenvolvimento de cancro do cólon na maioria das famílias FCCTX. Desta forma, a identificação de novos genes/variantes genéticas específicas que contribuam para a susceptibilidade para esta doença reveste-se de extrema importância.

O gene *FBXO11* foi associado a um mecanismo de interacção *antisense* com o gene *MSH6*, através de um estudo realizado por Duan et al., 2012. Anteriormente, o nosso grupo detectou tumores FCCTX com deleções na região promotora do gene *MSH6*. Deste modo, colocou-se a hipótese da ocorrência de um evento germinal no gene *FBXO11* e de um segundo evento (somático) corresponder às deleções observadas no gene *MSH6*. Os resultados da análise mutacional e das análises de expressão e *copy number* para o gene *FBXO11* não revelaram qualquer alteração que pudesse contribuir para a susceptibilidade para o FCCTX. No entanto, não se pode excluir a possível interacção destes dois genes num contexto de desenvolvimento de CCR esporádico.

A análise de um painel multigénico (*Trusight Cancer Illumina*) de 94 genes associados a um risco aumentado para cancro, sugeriu que mutações nos genes de reparação associados à HR podem contribuir para o risco aumentado de CCR e para a susceptibilidade para o FCCTX. O facto de mutações nestes genes terem sido identificadas apenas nos indivíduos index de famílias TSG⁺, tendo em conta a assinatura molecular observada nos tumores destes indivíduos, a qual pode ser explicada por uma deficiente reparação de DSB devido a defeitos no sistema de reparação associado à HR, reforça a possibilidade de as referidas mutações poderem explicar e/ou contribuir para o FCCTX nestas famílias. Por outro lado, estes resultados também permitiram confirmar que as duas entidades moleculares distintas (TSG⁺ e TSG⁻) deverão seguir vias diferentes no contexto da susceptibilidade para FCCTX. Mais ainda, em algumas famílias, mutações em dois genes distintos poderão contribuir em conjunto para o fenótipo apresentado. Este facto, vem reforçar a hipótese que tem vindo a ser proposta em vários estudos de que em muitos casos esta síndrome não seja monogénica, mas antes poligénica, resultando da interacção de vários genes de baixa penetrância. A detecção de mutações nestes genes em algumas famílias HNPCC ou que preenchem os CB, nestes últimos essencialmente nos critérios associados a história familiar de CCR, sugere que algumas famílias destes sub-grupos de doentes poderão ser potenciais FCCTX. Assim a análise destes genes poderá vir a ser útil para o manejo clínico destas famílias.

A sequenciação do exoma numa família informativa de FCCTX (TSG⁻), utilizou vários modelos de segregação onde se excluíram eventuais fenocópias, levando à selecção de variantes germinais possivelmente patogénicas. Tendo em conta a análise de segregação e a sua previsão de patogénecidade *in silico*, identificou-se mais do que uma variante capaz de contribuir para a susceptibilidade para esta síndrome e para o aumento do risco de desenvolvimento de carcinoma ou adenoma (variantes presentes nos genes *MTMR3*, *LGR6*, *NEK7*, *DUSP12*, *NPR2* e *TASIR1*). A

variante identificada no gene *MTMR3* (p.Arg645Trp) estava presente em indivíduos afectados quer com carcinoma quer com adenomas com a excepção do indivíduo L1094 que até ao momento é o que apresenta um fenótipo menos severo (presença de apenas dois adenomas) e não se pode excluir se não poderão ter sido desenvolvidos no contexto esporádico. Os resultados tanto da análise de expressão como da análise mutacional, assim como as funções que a proteína MTMR3 parece apresentar, indicam uma possibilidade deste gene ser um candidato para a susceptibilidade para FCCTX. à semelhança do observado na análise mutacional utilizando o painel multigénico, para a qual também foram constatados indivíduos com mutação em mais do que um gene, é possível que alguma(s) das restantes 5 variantes, resultantes da análise do exoma, possa também contribuir para o fenótipo na família, de acordo com um modelo do FCCTX como uma doença poligénica. Este facto confere uma maior dificuldade no esclarecimento da sua base genética, assim como na detecção da causa molecular subjacente à doença em cada família.

7. Referências

- Abbas, T., Mueller, A. C., Shibata, E., et al. 2013. CRL1-FBXO11 promotes Cdt2 ubiquitylation and degradation and regulates Pr-Set7/Set8-mediated cellular migration. *Mol Cell*, 49(6):1147–1158.
- Abida, WM., Nikolaev, A., Zhao, W., et al. 2007. FBXO11 promotes the Neddylation of p53 and inhibits its transcriptional activity. *J Biol Chem*, 282: 1797-1804.
- Armaghany, T., Wilson, J. D., Chu, Q., et al. 2012. Genetic alterations in colorectal cancer. *Gastrointest Cancer Res*, 5(1):19–27.
- Ashktorab, H., e Brim, H. 2014. DNA Methylation and Colorectal Cancer. *Curr Colorectal Cancer Rep*, 10(4):425–430.
- Bamshad, M. J., Ng, S. B., Bigham, A. W., et al. 2011. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*, 12(11):745–755.
- Basu, S., Haase, G., e Ben-Ze'ev, A. 2016. Wnt signaling in cancer stem cells and colon cancer metastasis. *F1000Res*, 5.
- Belo, Hélio. 2010. Papel Do Gene APC e Identificação de Regiões Cromossômicas Envolvidas No Síndrome Familiar de Câncer Do Cólon e Recto Do Tipo X. Tese de Mestrado. Escola Superior de Saúde Egas Moniz.
- Birchmeier, W. 2011. Stem cells: Orphan receptors find a home. *Nature*, 476(7360):287–288.
- Blondeau, F., Laporte, J., Bodin, S., et al. 2000. Myotubularin, a phosphatase deficient in myotubular myopathy, acts on phosphatidylinositol 3-kinase and phosphatidylinositol 3-phosphate pathway. *Hum Mol Genet*, 9(15):2223–2229.
- Boland, C.R. (2005). Evolution of the nomenclature for the hereditary colorectal cancer syndromes. *Fam Cancer*, 4(3):211–218.
- Boland, C.R., e Goel, A. 2010. Microsatellite instability in colorectal cancer. *Gastroenterology*, 138(6):2073–2087.
- Brandão, C., e Lage, J. 2015. Management of Patients with Hereditary Colorectal Cancer Syndromes. *GE Portuguese Journal of Gastroenterology*, 22(5):204–212.
- Breaks, D. N. A. D., Lu, H., Shamanna, R. A., et al. 2016. RECQL4 Promotes DNA End Resection in Repair of Double-Strand Breaks. *Cell Rep*, 16(1):161–173.
- Burt, R. 2007. Inheritance of Colorectal Cancer. *Drug Discov Today Dis Mech*, 4(4):293–300.

Cain, E., e Besser,A. 2013. Emerging Roles of Atypical Dual Specificity Phosphatases in Cancer. *Oncogene and Cancer - From Bench to Clinic*. InTech.

Chen, L., Huang, S., Lee, L., et al. 2003. WRN , the protein deficient in Werner syndrome , plays a critical structural role in optimizing DNA repair. *Aging Cell*, 2(4):191–199.

Chenevix-trench, G., Liu, C., Srihari, S., et al. 2014. A fine-scale dissection of the DNA double-strand break repair machinery and its implications for breast cancer therapy. *Nucleic Acids Res*, 42(10):6106–6127.

Choi, J. H., Lee, J., Choi, I. J., et al. 2016. Variations in TAS1R taste receptor gene family modify food intake and gastric cancer risk in a Korean population. *Mol Nutr Food Res*,60(11): 2–14.

Church, J., Wood, L. D., Parsons, D. W., et al. 2016. Molecular genetics of colorectal Cancer. *Seminars in Colon and Rectal Surgery*, 27(4):172-175.

Colussi, D., Brandi, G., Bazzoli, F., et al. 2013. Molecular pathways involved in colorectal cancer: implications for disease behavior and prevention. *Int J Mol Sci*,14(8):16365-16385.

Daftuar, L., Zhu, Y., Jacq, X., et al. 2013. Ribosomal Proteins RPL37, RPS15 and RPS20 Regulate the Mdm2-p53-MdmX Network. *PLoS ONE*, 8(7): e68667.

Dallosso, A. R., Jones, S., Azzopardi, D., et al. 2009. The APC variant p.Glu1317Gln predisposes to colorectal adenomas by a novel mechanism of relaxing the target for tumorigenic somatic APC mutations. *Hum Mutat*, 30(10):1412–1418.

Dominguez-Valentin, M., Therkildsen, C., Da Silva, S., et al. 2015. Familial colorectal cancer type X: genetic profiles and phenotypic features. *Mod Pathol*, 28(1):30–36.

Duan, S., Cermak, L., Pagan, J. K.,et al. 2012. FBXO11 targets BCL6 for degradation and is inactivated in diffuse large B-cell lymphomas. *Nature*, 481(7379):90–93.

Duarte, M. 2015. Envolvimento de variantes genéticas específicas na suscetibilidade para o risco aumentado de cancro do cólon e reto e na agressividade tumoral. Tese de Mestrado. Escola Superior de Saúde Egas Moniz.

Espina, C., Céspedes, M. V., García-Cabezas, M. A., et al. 2008. A critical role for Rac1 in tumor progression of human colorectal adenocarcinoma cells. *Am J Pathol*, 172(1):156–166.

Ferreira, S., Lage, P., Sousa, R., 2009. Familial colorectal cancer type X: clinical, pathological and molecular characterization. *Acta MedPort*, 22(3), 207–214.

- Francisco, I., Albuquerque, C., Lage, P., et al. 2011. Familial colorectal cancer type X syndrome: two distinct molecular entities?. *Fam Cancer*, 10(4):623–631.
- Fry, A. M., O'Regan, L., Sabir, S. R., et al. 2012. Cell cycle regulation by the NEK family of protein kinases. *J Cell Sci*, 125(19):4423–4433.
- Gimeno-García, A. Z., Hernández-Álvarez-de-Buylia, N., Nicolás-Pérez, D., et al. 2015. Colorectal cancer screening in the familial risk population: Is colonoscopy still the strategy of choice?. *Gastroenterol Hepatol*, 39(5):352–60.
- Girard, B. J., Daniel, A. R., Lange, C. A., et al. 2015. PELP1: a review of PELP1 interactions, signaling and biology. *Mol Cell Endocrinol*, 382(1):612–626.
- Hahn, M. A., Dickson, K., Jackson, S., et al. 2012. The tumor suppressor CDC73 interacts with the ring finger proteins RNF20 and RNF40 and is required for the maintenance of histone 2B monoubiquitination. *Mol cell*, 21(3):559–568.
- Hannig, G., Tchernychev, B., Kurtz, C. B., et al. 2014. Guanylate cyclase-C/cGMP: an emerging pathway in the regulation of visceral pain. *Front Mol Neurosci*, 7:31.
- Hao, F., Itoh, T., Morita, E., et al. 2016. The PtdIns3-phosphatase MTMR3 interacts with mTORC1 and suppresses its activity. *FEBS Lett*, 590(1):161–73.
- Hawkins, A. T., Wise, P. E., Bisgaard, M. L., et al. 2016. Colon cancer in hereditary syndromes. *Gastroenterology*, 138(6):121–125.
- Heo, J., Eki, R., e Abbas, T. 2016. Deregulation of F-box proteins and its consequence on cancer development, progression and metastasis. *Semin Cancer Biol*, 36:33–51.
- Hnia, K., Vaccari, I., Bolino, A., et al. 2012. Myotubularin phosphoinositide phosphatases: cellular functions and disease pathophysiology. *Trends Mol Med*, 18(6):317–327.
- Huber, A., Findeis-Hosey, J. J., e Whitney-Miller, C. L. 2013. Hereditary Gastrointestinal Polyposis Syndromes: A Review Including Newly Identified Syndromes. *J Gastroint Dig Syst*, 3:155.
- Jaspersion, K. W., Tuohy, T. M., Neklason, D. W., et al. 2010. Hereditary and familial colon cancer. *Gastroenterology*, 138(6):2044–2058.
- Kerr, D. 2003. Clinical development of gene therapy for colorectal cancer. *Nat Rev Cancer*, 3(8):615–622.

Kim, T.-H., Leslie, P., e Zhang, Y. 2014. Ribosomal proteins as unrevealed caretakers for cellular stress and genomic instability. *Oncotarget*, 5(4):860–871.

Laczmanska, I., Gil, J., Karpinski, P., et al. 2007. Polymorphism in nucleotide excision repair gene XPC correlates with bleomycin-induced chromosomal aberrations. *Environ Mol Mutagen*, 48(8):666–671.

Lanza, G., Gafà, R., Maestri, I., et al. 2002. Immunohistochemical Pattern of MLH1/MSH2 Expression Is Related to Clinical and Pathological Features in Colorectal Adenocarcinomas with Microsatellite Instability. *Mod Pathol*, 15(7):741–749.

Leedham, S. J., e Chetty, R. 2016. Wnt disruption in colorectal polyps - the traditional serrated adenoma enters the fray. *J Pathol*, 239(4):387–90.

Ligtenberg, M. J., Kuiper, R. D., Chan, T. L., et al. 2009. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nat Genet*, 41(1):112–117.

Lindor, N. M., Rabe, K., Petersen, G. M., et al. 2005. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. *JAMA*, 293(16):1979–1985.

Liu, L., e Gerson, S. L. 2006) Targeted Modulation of MGMT: Clinical Implications. *Clin Cancer Res*. 12(2):328–331.

Lorenzo, O., Urbé, S., e Clague, M. J. 2005. Analysis of phosphoinositide binding domain properties within the myotubularin-related protein MTMR3. *J Cell Sci*, 118(Pt 9):2005–2012.

Luo, H., Cowen, L., Yu, G., et al. 2016. SMG7 is a critical regulator of p53 stability and function in DNA damage stress response. *Cell Discov*, 2:15042.

Matos, P., e Jordan, P. 2008. Increased Rac1b expression sustains colorectal tumor cell survival. *Molecular Cancer Research : Mol Cancer Res*, 6(7):1178–1184.

Michl, J., Zimmer, J., e Tarsounas, M. 2016. Interplay between Fanconi anemia and homologous recombination pathways in genome integrity. *EMBO J*, 35(9):909–923.

Miller, S. A., Dykes, D. D., e Polesky, H. F. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*, 16(3):1215.

Miura, K., Namba, N., Fujiwara, M., et al. 2012. An Overgrowth Disorder Associated with Excessive

Production of cGMP Due to a Gain-of-Function Mutation of the Natriuretic Peptide Receptor 2 Gene. *PloS One*, 7(8):1–10.

Mokarram, P., Kumar, K., Brim, H., et al. 2009. Distinct high-profile methylated genes in colorectal cancer. *PloS One*, 4(9):e7012.

Morriissy, A. S., Griffith, M., e Marra, M. A. 2011. Extensive relationship between antisense transcription and alternative splicing in the human genome. *Genome Res*, 21(8):1203–12.

Nieminen, T. T., Abdel-Rahman, W. M., Ristimäki, A., et al. 2011. BMPR1A mutations in hereditary nonpolyposis colorectal cancer without mismatch repair deficiency. *Gastroenterology*, 141(1):e23–6.

Nieminen, T. T., O'Donohue, M.-F., Wu, Y., et al. 2014. Germline mutation of RPS20, encoding a ribosomal protein, causes predisposition to hereditary nonpolyposis colorectal carcinoma without DNA mismatch repair deficiency. *Gastroenterology*, 147(3):595–598.

Ogino, S., Nosh, K., Kirkner, G. J., et al. 2009. CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer. *Gut*, 58(1), 90–96.

Oji, Y., Tatsumi, N., Kobayashi, J., et al. 2015. Wilms ' Tumor Gene WT1 Promotes Homologous Recombination-Mediated DNA Damage Repair. *Mol Cell Carcinog*, 54(12):1758–1771.

Ong, B. A., Vega, K. J., e Houchen, C. W. 2014. Intestinal stem cells and the colorectal cancer microenvironment. *World J Gastroenterol*, 20(8):1898–909.

Oppelt, A., Haugsten, E. M., Zech, T., et al. 2014. PIKfyve, MTMR3 and their product PtdIns5P regulate cancer cell migration and invasion through activation of Rac1. *Biochem J*, 461(3):383–90.

Palles, C., Cazier, J.-B., Howarth, K. M., et al. 2013. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet*, 45(2):136–144.

Patterson, K. I., Brummer, T., O'Brien, P. M., et al. 2009. Dual-specificity phosphatases: critical regulators with diverse cellular targets. *Biochem J*, 418(3):475–89.

Peng, G., e Lin, S.-Y. 2011. Exploiting the homologous recombination DNA repair network for targeted cancer therapy. *World J Clin Oncol*, 2(2):73–79.

Pereira, C. 2013. Estudo do gene NRIP1 e de novos *loci* de susceptibilidade para o cancro do cólon e recto familiar do tipo X. Escola Superior de Saúde Egas Moniz.

- Pino, M. S., e Chung, D. C. 2010. The chromosomal instability pathway in colon cancer. *Gastroenterology*, 138(6):2059–2072.
- Plant, P. J., Correa, J., Goldenberg, N., et al. 2009. The inositol phosphatase MTMR4 is a novel target of the ubiquitin ligase Nedd4. *Biochem J*, 419(1):57–63.
- Rabbani, B., Tekin, M., e Mahdieh, N. 2014. The promise of whole-exome sequencing in medical genetics. *J Hum Genet*, 59(1):5–15.
- Randle, S. J., e Laman, H. 2016. F-box protein interactions with the hallmark pathways in cancer. *Semin Cancer Biol*, 36:3–17.
- Regan, L. O., e Fry, A. M. 2009. The Nek6 and Nek7 Protein Kinases Are Required for Robust Mitotic Spindle Formation and Cytokinesis. *Mol Cell Biol*, 29(14):3975-3990
- Renato, J., Cussiol, R., Cussiol, J. R., et al. 2016. Slx4 scaffolding in homologous recombination and checkpoint control : lessons from yeast. *Chromossoma*.
- Roper, J., e Hung, K. E. 2013. Molecular Mechanisms of Colorectal Carcinogenesis. *Molecular Pathogenesis of Colorectal Cancer* (pp. 25–65).
- Saramago, A. 2014. Estudo de mutações germinais em genes de suscetibilidade para o Cancro do Colón e Reto Familiar do tipo X. Tese de Mestrado. Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.
- Schepers, A., e Clevers, H. 2012. Wnt signaling, stem cells, and cancer of the gastrointestinal tract. *Cold Spring Harb Perspect Biol*, 4(4): a007989.
- Schulz, E., Klampfl, P., Holzapfel, S., et al. 2014. Germline variants in the SEMA4A gene predispose to familial colorectal cancer type X. *Nat Commun*, 5:5191.
- Scott, R. J., Crooks, R., Rose, L., et al. 2004. Germline Missense Changes in the APC Gene and Their Relationship to Disease. *Hered Cancer Clin Pract*, 2(2):81–91.
- Shiovitz, S., Copeland, W. K., Passarelli, M. N., et al. 2014. Characterisation of familial colorectal cancer Type X, Lynch syndrome, and non-familial colorectal cancer. *Br J Cancer*, 111(3), 598–602.
- Silva, P., Albuquerque, C., Lage, P.,. 2016. Serrated polyposis associated with a family history of colorectal cancer and / or polyps : The preferential location of polyps in the colon and rectum defines two molecular entities. *Int J Mol Med*, 38(3): 687–702.
- Sjöblom, T., Jones, S., Wood, L. D., et al. 2006. The consensus coding sequences of human breast and colorectal cancers. *Science*, 314(5797): 268–74.

- Skaar, J. R., Pagan, J. K., e Pagano, M. 2013. Mechanisms and function of substrate recruitment by F-box proteins. *Nat Rev Mol Cell Biol*, 14(6):369–381.
- Stoffel, E. M., e Kastrinos, F. 2014. Familial colorectal cancer, beyond Lynch syndrome. *Clin Gastroenterol Hepatol*, 12(7), 1059–1068.
- Taguchi-Atarashi, N., Hamasaki, M., Matsunaga, K., et al. 2010. Modulation of local PtdIns3P levels by the PI phosphatase MTMR3 regulates constitutive autophagy. *Traffic*, 11(4):468–478.
- Tamura, N., e Garbers, D. L. 2003. Regulation of the Guanylyl Cyclase-B Receptor by Alternative Splicing. *J Biol Chem*, 278(49):48880–48889.
- Taylor, G. S., Maehama, T., e Dixon, J. E. 2000. Myotubularin, a protein tyrosine phosphatase mutated in myotubular myopathy, dephosphorylates the lipid second messenger, phosphatidylinositol 3-phosphate. *Proc Natl Acad Sci U S A*, 97(16):8910–8915.
- Tomlinson, I. 2015. An update on the molecular pathology of the intestinal polyposis syndromes. *Diagnostic Histopathology*, 21: (pp. 147–151).
- Umar, A., Risinger, J. I., Hawk, E. T., et al. 2004. Testing guidelines for hereditary non-polyposis colorectal cancer. *Nat Rev Cancer*, 4(2):153–8.
- Wang, Z., Liu, P., Inuzuka, H., et al. 2014. Roles of F-box proteins in cancer. *Nat Rev Cancer*, 14(4):233–247.
- Warr, A., Robert, C., Hume, D., et al. 2015. Exome Sequencing: Current and Future Perspectives. *G3 (Bethesda, Md.)*, 5(8):1543–1550.
- Watson, A. J. M., e Collins, P. D. 2011. Colon Cancer: A Civilization Disorder. *Dig Dis*, 29(2):222–228.
- Wauson, E. M., Zaganjor, E., e Cobb, M. H. 2013. Amino acid regulation of autophagy through the GPCR TAS1R1-TAS1R3. *Autophagy*, 9(3):418–419.
- Wauson, E. M., Zaganjor, E., Lee, A. Y., et al. 2012. The G Protein-Coupled Taste Receptor T1R1/T1R3 Regulates mTORC1 and Autophagy. *Mol Cell*, 47(6):851–862.
- Weiss, M. J., Santos, C. O., e De, W. 2009. Chaperoning erythropoiesis- Review article Chaperoning erythropoiesis. *Blood*, 113(10):2136–2144.
- Woods, M. O., Younghusband, H. B., Parfrey, P. S., et al. 2010. The genetic basis of colorectal cancer in a population-based incident cohort with a high rate of familial disease. *Gut*, 59(10): 1369–1377.

Yan, P., Klingbiel, D., Saridaki, Z., et al. 2016. Reduced Expression of SMAD4 Is Associated with Poor Survival in Colon Cancer. *Clin Cancer Res*, 22(12):3037-3047.

Zheng, B., Yu, X., Chai, R., et al. 2014. Myotubularin-related phosphatase 3 promotes growth of colorectal cancer cells. *The Scientific World Journal*, 703804.

Zheng, H., Shen, M., Zha, Y.-L., et al. 2014. PKD1 phosphorylation-dependent degradation of SNAIL by SCF-FBXO11 regulates epithelial-mesenchymal transition and metastasis. *Cancer Cell*, 26(3):358–73.

Zhu, B., Zheng, Y., Pham, A., et al. 2005. Monoubiquitination of Human Histone H2B : The Factors Involved and Their Roles in HOX Gene Regulation. *Mol Cell*, 20(4):601–611.

8. Anexos

Anexo A: Sequência e características dos *primers* utilizados neste estudo

- I. Sequências dos *primers* e tamanho do fragmento esperado utilizados para amplificação dos vários exões do gene *FBXO11* (Adaptado de: Duan et al., 2012)

Gene	Exão/ exões amplificados	Sequência do <i>primer</i>		T _m (C)	Tamanho do fragmento(bp)
<i>FBXO11</i>	1 a 2	Fw	CTCCCTCCCGAATTGAAG	57,01	609
		Rev	AAAGTCAGAGGGAGAGGTCAGG		
	3	Fw	TGTGTTACCCATGAAACCCAC	57,9	369
		Rev	CCGACTTTCCTACCATGTTTAGC	60,18	
	4	Fw	TCTAGCCTGGATGACAGTGAGAC	58,51	398
		Rev	CGCCCAGCCTAAACTTATTTTC	58,92	
	5 a 6	Fw	GAAACCCTGTCTTTGTTTCTGG	59,92	617
		Rev	TCTTCATTCTACTTTACCAGCAG	58	
	7 a 8	Fw	AGCAAAGGCTGATGATGAAA	55,75	583
		Rev	GGACATGCTCCTTGACCACT	56,87	
	9 a 10	Fw	ACACCACAATGCACACCACT	56,12	714
		Rev	GAACGCTATCACCTCTACATGG	56,76	
	11	Fw	GCCCAGCCTTAAATGTTTCTAATAC	60,58	503
		Rev	CAATAGCTATGGCTCATTGAGATG	58,97	
	12	Fw	TGAAGAAATTGAGGCCTTAGAAGT	59,04	315
		Rev	TCAGTGCTTCCACTTGGGTAG	57,72	
	13	Fw	GCCTGCAGCTCTCTGCATC	58,88	413
		Rev	TTAAGATACTGGCAGGGAGAAGAAG	60,85	
	14 a 15	Fw	TGCCATTACCTCCTTACTCGG	59,51	625
		Rev	CCAGTGGCTTCTGTCCTCAC	57,49	
	16	Fw	TTCTCAGGGCATCTTGGACTC	59,06	341
		Rev	GTGATCCACCCATCTCGG	55,63	
	17	Fw	CCACTGCACACTCCAGCC	57,1	410
		Rev	GGCACTAATCTCCTAAACCAGTTC	58,75	
	18	Fw	TGGAGATGGCAGATTATTGGTC	59,16	462
		Rev	CATTGAGCCACTTCAGCACAC	58,82	
	19	Fw	ACTGTTGGTGGTTGAAATAGGTATC	58,83	425
		Rev	TCCAACCATTAGAACCAAAGC	58,91	
	20	Fw	GCTTTGGTTCTAAATGGTTGGAG	60,14	481
		Rev	ACCACGCCTACCCACAGTTAC	59,06	
	21 a 22	Fw	GGTCTCTTAAAGCCAAAGGTC	59,2	610
		Rev	ACCCAGCTTTGAGATCCTGAG	58,37	

- II. Sequências dos *primers* utilizados na reacção de sequenciação, para alguns exões do gene *FBXO11*, em que estes são diferentes dos utilizados para a amplificação por PCR descritos na tabela I do anexo A (Adaptado de: Duan et al., 2012)

Gene	Exão / exões amplificados	Sequência do primer	
<i>FBXO11</i>	21 a 22	Fw	GGTCCTCTTAAAGCCAAAGGTC
	3	Fw	CATTAGATAAAATGAAAAGTCTG
	4	Fw	TCTTACTCAAGTTTTTGAGGC
	9 a 10	Fw	GACTATATATAACAGATCATG
		Rev	CATTATCCTCATATATTCCC
	14	Rev	AAAGCTTTTCAAGGGACAAG
	17	Fw	AAATATACCTTCTAGATAACC

- III. Sequências dos *primers* utilizados para a análise de expressão e de *copy-number* do gene *FBXO11*, por qPCR, e localização cromossómica dos mesmos (Duan et al., 2012)

Estudo	Gene	Sequência do primer		Localização cromossómica (tamanho em bp)
Análise de Expressão	<i>FBXO11</i>	Fw	GATGGACGAGGCCTTATTGA	Chr 2:47820397-47822254 (97)
		Rev	TGTTATGCCGAACAATTGGA	
	<i>SDHA</i>	Fw	TACAAGGTGCGGATTGATGA	Chr 5: 251463-254476 (90)
		Rev	GGTGTGCTTCTCCAGTGCT	
Copy- Number	<i>FBXO11_probe1</i>	Fw	CTCCATGACCACCACTCCTT	Chr 2: 26965666-26965733 (97)
		Rev	ATGACACAGCCCTCATCTCC	
	<i>FBXO11_probe2</i>	Fw	ACTTGGCCAATTTCAACCAG	Chr 2 : 26864659-26864712(54)
		Rev	TCAGACCCCAATTTTCAGAG	
	<i>CHR12_probe1</i>	Fw	GCTTGCCCTGTCCAGTTAAT	Chr 12: 6585217-6585317(101)
		Rev	TAGCTCAGCTGCACCCTTA	

- IV. Sequências e temperaturas de *annealing* dos *primers* utilizados para confirmação das variantes identificadas pela análise do painel multigénico *Trusight Cancer* (Illumina) nas amostras das 15 famílias FCCTX (informações obtidas com o software *NetPrimer*: (<http://www.premierbiosoft.com/netprimer/>) e *UCSC In-Silico PCR* (<https://genome.ucsc.edu/cgi-bin/hgPcr>))

Sub-grupo	Amostra	Gene	Variante	Sequência do primer		Tm (C)	Tamanho do fragmento (bp)
TSG+	L5	<i>RECQL4</i>	c.2761G>A, p.Glu921Lys	Fw	CACTCCCAATACAGCTTACCG	57	290
	L48	<i>CDC73</i>		Rev	CTCCAACCTCGTCTCCAAT		
	L108	<i>FANCA</i>	c.691T>C, p.Trp231Arg	Fw	GCCTCCCGAATGTAGCAGTTT	64	503
				Rev	AATGTTGGGGAGGGCTTTTAC		
	L484	<i>WRN</i>	c.1007-3C>A	Fw	CGCATTGATGGTAGTGGTCC	62	285
				Rev	TACCACATCCACTCACCTG		
	L984	<i>WT1</i>	c.435G>T, p.Trp145Cys	Fw	GCAGGAGTGGCATTCAAAC	58	351
				Rev	GGCAACATCTGTCAACTCC		
	L1038	<i>FANCE</i>	c.1109G>C, p.Arg370Pro	Fw	CACCTGCCACCCCTTCTTT	60	494
				Rev	CTCTACCACTCTGCTCTGC		
	L1038	<i>FANCE</i>	c.1316G>T, p.Gly439Val	Fw	AGGCTGGGCATTCTGTTACC	62	452
				Rev	TACAACCTGCCAGGGAGGTCA		
	L1200	<i>SLX4</i>	c.4301C>T, p.Ser1434Leu	Fw	CTTCTGAACCAGACCCCA	60	411
				Rev	TGCTCTCCCCGTCCCAAA		
	L1200	<i>XPC</i>	c.860T>G, p.Phe287Cys	Fw	TGCCCTGACCTCTGACACAA	62	248
				Rev	ACCTGGAATGGCATCTGGCA		

- III. Sequências e temperaturas de *annealing* dos *primers* utilizados para segregação das variantes identificadas pela análise de exoma na família L56 (informações obtidas com o software *NetPrimer*: (<http://www.premierbiosoft.com/netprimer/>) e *UCSC In-Silico PCR* (<https://genome.ucsc.edu/cgi-bin/hgPcr>))

Output	Gene	Caracterização da variante em estudo	Primer		Tm (C)	Tamanho do fragmento(bp)
4	C16orf96	c.611C>T ;p.Ser204Phe	Frw	GCAAGGAGGAGAACACAGC	60	219
			Rev	TTCCTGCTAACCTCCCCTT	58	
	C2orf70	c.232C>G;p.Pro78Ala	Frw	CCACAGCATCCTCCACTT	56	557
			Rev	AGTCCTGAGCCTCCTTAC		
	DCHS1	c.97_100delTGC; p.Cys34_Cys35ins	Frw	CATTGTGCCTTCCTGCCCT	60	385
			Rev	AGCCACTCGCACTGTAAC	58	
	DUSP12	c.1_4insTTG; p.Leu2_Leu3ins;	Frw	ATCAACTTTGGGGGTAGGG	58	331
			Rev	CGAAATACAATCCTGGCTGC	60	
	GP1BA	c.775G>A;p. Val259Met	Frw	CGACACCCTTCTCTCCAA	57,86	494
			Rev	ATGTGGTCTGCTCCTTAGTG	52,16	
	MAGI1	c.1263_1264insCAG; p. Gln421_Thr422insGln	Frw	TGCTGGTGGGATGAACTGT	58	306
			Rev	TCAGTGTGGCAAGAGAGGA		
	MAML2	c.1821_1829delACAGCAGCA;p.Gln472_Gln474del	Frw	TCAAGCGAACCAGCAGATG	57,56	493
			Rev	GCTAAATGGTCTCTGGGAAC	53,8	
	MTMR3	c.1933C>T;p.Arg645Trp	Frw	CCAAGATTTTCATTTCCCCC	58,57	406
			Rev	TTCCTCCTGCTCTCCTTTTC	56,12	
	NPR2	c.2644G>A;p.Val882Ile	Frw	TTGGCTTCACAGCATTGTCA	58	346
			Rev	AAAGCGGGTCTCTATTAAAGG	55,58	
	OR3A1	c.392C>T;p.Arg131Gln	Frw	CAGCGTCACTGTTCCATCA	58	385
			Rev	CACAGCAAAAAGCAGCAGC		
	OR56A4	c.464T>C; p.Q155R	Frw	CTAACACCACCTCCTGAT	58	359
			Rev	GGGAACAGGAAGAGAAACA	56	
	PROSER2	c.G607A; p.Glu203Lys	Frw	AAACAAGATGCTGAGACTCCT	53,15	362
			Rev	TTGGTGACGATGATGTTGCT	56,69	
	RNF207	c.1615C>T; p.Arg2695Cys	Frw	TGGTGCCTGGCTTGAGTAA	57,12	300
			Rev	TCAGCCTCCCGAAGTGTTA	56,21	
	SMG7	c.1757C>A; p.Pro657His	Frw	ACTCCCTCTACCTACCTGT	58	380
			Rev	CTGGCTTGAGACTGTGCT		
	TNIP3	c.854T>G; p.Glu285Ala	Frw	AGGGATGGATGACTTGCTC	58	431
			Rev	GACTTCATCCCAGGTGCTG	60	

<i>Output</i>	<i>Gene</i>	<i>Caracterização da variante em estudo</i>	<i>Primer</i>		<i>Tm (C)</i>	<i>Tamanho do fragmento(bp)</i>
3	<i>AKR1C3</i>	c.623C>T; p.Ser208Leu	Fw	GAGCTTCTGTTCAGAAATGGC	56,74	397
			Rev	AGGAAACTGAGGATGCTCTGA	56,34	
	<i>ANKRD12</i>	c.297A>G; p.Thr99=	Fw	CTGCATGGTTGAATTTCCCA	59,04	446
			Rev	AAATGTGGTGTACCCACGTA	58,33	
	<i>ARHGAP22</i>	c.963G>A; p.Asn321=	Fw	GCTTCTTTGATGACTGCTGCT	57,33	294
			Rev	AGATGGAGATGTGCCACCTGA	59,93	
	<i>BBOX1</i>	c.639+5_639+6insG	Fw	TTTCCTCTGGCTTAGTGACAA	55,25	558
			Rev	CCATGTGACTTTGAGGGAAA	55,56	
	<i>CACNAIS*</i>	c.3607C>T; p.D1203N	Fw	TTTCCCATCCTTTTCCTTCC	58,27	243
			Rev	AGATGGAATGACAAGTCACGG	57,17	
	<i>BRD4</i>	c.3810C>G; p.Glu1270Asp	Fw	AAATCAAGAATCGGGCTCCT	57,99	564
			Rev	TCAGGAGCTAATCCTTAGACCA	56,29	
	<i>CEP250</i>	c.4208A>C; p.Gln1403Pro	Fw	TTGCCCAAGTCCTTATCACTG	57,8	439
			Rev	TCTCCAGTTCCTGGATTGTC	56,42	
	<i>CLIP2</i>	c.2238G>A; p.Ala746=	Fw	AGTGATGGAGGGCATCAAGAT	58,33	547
			Rev	TGCAGAACAGAGGTCAGCAAA	59,18	
	<i>GALNT8</i>	c.988G>A; p.Glu330Lys	Fw	ATCTTGGCTCGGATTCAGGA	59,42	239
			Rev	TTCCTGCCTTCCACATGCTA	62,88	
	<i>KIAA1462</i>	c.1660T>C; p.Thr554Ala	Fw	ATCCACAGAGCTTAATACCCC	55,39	400
			Rev	TGATCAGCACTGGGCTTTAAG	58,16	
	<i>LGR6</i>	c.1868G>A; p.Arg623Gln	Fw	CAACACCTTGACTGGCATTTC	57,95	404
			Rev	AGGAGTTCATCATCACCAGGG	58,57	
	<i>MAP4</i>	c. 6422G>A; p.Ser2141Phe	Fw	GGGTGGAGGTTAACAACACT	55,82	307
			Rev	TTCTATCTACATGGGCACCTG	55,29	
	<i>MAP6</i>	c.1716A>G; p.Asn572=	Fw	TATCTCAGCACCAGTCAAGGA	55,48	256
			Rev	TTGATCCTTGACAGGTGCTGA	58,68	
	<i>NEK7</i>	c.103C>G; p.Arg35Gly	Fw	GAGTGAATGGATGTCTGCAAA	55,71	359
			Rev	AGGCTATCAGTACCCAATCAT	53,23	
	<i>PARD3</i>	c.2540G>A; p.Ala847Val	Fw	TGGGTCTGAAGAAGTCAAGCT	56,49	427
			Rev	TGCAGTATCACAGGACATGTG	54,5	
	<i>PELP1</i>	c.894C>T; p.Ser298=	Fw	GCCATCTTTGCTCATGACACA	59,15	232
			Rev	AGCAGGCACACTTGGTCAAA	58,82	
	<i>PFKP</i>	c.339C>T; p.Asn113=	Fw	ATTAGCTGGTGTCTTTCCAG	56,27	413
			Rev	CCTGGCACATGGTAAGGTAA	57,88	
	<i>PTPN22</i>	c.197-4A>G	Fw	TCACTTGTTCACTCATGAGCA	54,29	320
			Rev	CTTGCGTTCAAGTGATCTTTC	55,21	
	<i>PTPN23</i>	c.3988G>A; p.Val1330Met	Fw	TCTTGGCTTATCTGTCCCTCA	57,35	260
			Rev	AAGGACTGGACAGCACTTAGA	54,06	
	<i>RASL11B</i>	c.577G>A; p.Ala193Thr	Fw	AAACAGGTTGACCCTCAGCTT	58,05	353
			Rev	GCCAATATTCAACGTTCTCTG	55,7	
	<i>RNF217</i>	c.6+3A>G	Fw	TCCATTGCCAGCACTTGAGTA	59,09	400
			Rev	AGAATAGTCAGCTGCCAGGCA	60,05	
	<i>SULTA1</i>	c.-44C>A	Fw	AGCAAACCTCAGTCGTGGCTTT	59,23	391
			Rev	CAGGGTTGTCTGAAATGGGAT	58,76	
	<i>TM4SF5</i>	c.T179C; p.Val60Ala	Fw	AGGATAAAGTAAGGTGAAGCC	52,8	310
			Rev	ATTACAGGCATGAGACACCA	53,18	
	<i>TP63</i>	c.-72_-71insAGAG	Fw	AGTTGATGGATTGGACAGGTA	54,32	308
			Rev	GCATCTCTAAATGGAGTGCTA	52,53	
	<i>WDR37</i>	c.C531T_p.Ala177=	Fw	GTTTCACAAGACTCCCCATCT	55,52	327
			Rev	CACTAAGCATCACCAAGGGTT	56,92	
	<i>ZNF75A</i>	c.A931G	Fw	ATAGTGCCGGAAAATCTCCT	55,74	361
			Rev	TTGGTGCCATTTTCCCACT	58,37	

Análise Bioinformática anterior	Gene	Caracterização da variante em estudo	Primer		Tm (C)	Tamanho do fragmento(bp)
	CYB5R2	c.55C>T;p.Pro19Ser	Frw	TTGGTGTGGGGAGGGTGTT	60	315
			Rev	GCCATGAGATGACTGTCTG	58	
	DNAH3	c.10915C>A;p.Leu3639Ile	Frw	TTTCCAGTCAGCATTCTCCA	56,16	231
			Rev	GCACTGTCATCGCTTACCTA		
	GIMAP1*	c.908T>G;p.Val303Gly	Frw	TCCAACGAGGTGTATGAGC	58	559
			Rev	TGCGTTTGTCCCCACTACA		
	OBSCN	c.8365G>C;p.Gly2789Arg	Frw	AAGACAATGGAGGTGCTGGA	57,08	252
			Rev	CTGCCTACCTCTGACAATGA	53,08	

Legenda: *Condições de PCR ainda não otimizadas

III. Sequência dos *primers* utilizado para amplificação por PCR da variante do gene *MTMR3* ao nível somático

Fragmento (tamanho)	Variante	Sequência do primer	
<i>MTMR3</i> (129 bp)	p.Arg645Trp	Fw	ACAACACAGTGCCTCTGGCCA
		Rev	TGTCGGCAGAAAGGGGATCCT

IV. Sequências dos *primers* utilizados para a análise de expressão dos genes *MTMR3*, *RAC1* e *RAC1b* por qPCR, e localização cromossômica dos mesmos (os *primers* identificados com * encontram-se descritos em Duan et al., 2012)

Gene	Sequência do primer		Localização cromossômica (tamanho do fragmento em bp)
<i>GADPH</i> *	Fw	TGCACCACCAACTGCTTAGC	Chr 12: 6537319-6537598 (87)
	Rev	GGCATGGACTGTGGTCATGAG	
<i>MTMR3</i>	Fw	CGTGGAGTTCGATTCTTCTCTG	Chr22: 30007309-30009028 (154)
	Rev	TGGGTAATACTCTGGGCATTTCG	
<i>RAC1</i>	Fw	GGGCAAAGACAAGCCGATTG	Chr7: 6398697-6400168 (65)
	Rev	ATGATGCAGGACTCACAAGGG	
<i>RAC1b</i>	Fw	GGAGACGGAGCTGTAGGTAAA	Chr 7:6374763-6391983 (140)
	Rev	CATAAGCCCAGATTACCCGG	

Anexo B: Condições de PCR utilizadas para amplificação das variantes estudadas

I. Condições de PCR utilizada na amplificação dos exões do gene *FBXO11*

Gene	Exão amplificado	MgCl ₂ (μl)	Primer (μl)	Temperatura (°C)	Nº ciclos
<i>FBXO11</i>	1 a 2	1,5	0,2	68	35
	3	1,5	0,2	64,5	35
	4 *	3,25	0,3	68	35
	5 a 6	1,25	0,2	65	35
	7 a 8**	1,75	0,3	67	35
	9 a 10	0,75	0,3	67	35
	11	2,2	0,2	70	35
	12*	2,5	0,2	68	35
	13	1,75	0,2	67	35
	14 a 15	1,75	0,3	70	35
	16* /***	5,75	0,2	66,5	30
	17*	0,75	0,2	63	32
	18	2	0,3	69	35
	19	1,75	0,3	66,5	35
	20	1,25	0,3	65	35
	21 a 22	1,5	0,3	69,5	35

Legenda: * Utilização dos reagentes do kit *AmpliTaq Gold®* (Thermofisher); ** Adição de 3% de DMSO***Adição de 1,25 μl *GC rich resolution PCR System* (Roche);

II. Condições de qPCR e amostras utilizadas para as análises de expressão e de *copy number* do gene *FBXO11*

Análise	Gene	Amostras	Concentração dos <i>primers</i>	Concentração da amostra	Kit utilizado	Eficiência da recta
Análise de expressão	<i>FBXO11</i>	cDNA: L1484, L1624, L1846, L1855, L1956, L1871, L1807, L1897, L1811, L2120, L1363, L1547, PS, L1570, L1571	10 pmol/ul	15 ng/ul	<i>Applied Biosystems</i>	2,04
	<i>SDHA</i>		5 pmol/ul			1,91
Copy-number	<i>FBXO11 probe 1</i>	DNA: L5, L48, L108, L295, L926, L408, L674, L451, L720, E436, L467, L484, L499, L701, L984, L1038, L1200, L1121, L1152, L1484, L1635, L1624, L1846, L1855, L1956	5 pmol/ul	10 ng/ul	<i>Applied Biosystems</i>	1,99
	<i>FBXO11 probe 2</i>		10 pmol/ul			2,04
	<i>Chr 12</i>		5 pmol/ul			2,09

III. Condições de PCR utilizada na amplificação das variantes identificadas pela análise do painel multigênico *Trusight Cancer (Illumina)* nas amostras das 15 famílias FCCTX

Sub-grupo	Amostra	Gene	Variante	Tm (C)	MgCl2 (µl)
TSG+	L5	<i>RECQL4</i>	c.2761G>A, p.Glu921Lys	58	0,5
	L48	<i>CDC73</i>	c.691T>C, p.Trp231Arg	59	0,5
	L108	<i>FANCA</i>	c.1007-3C>A	67,5	1
	L484	<i>WRN</i>	c.435G>T, p.Trp145Cys	60	1
	L984	<i>WT1</i>	c.1109G>C, p.Arg370Pro	70	1
	L1038	<i>FANCE</i>	c.1316G>T, p.Gly439Val	64	0,5
		<i>SLX4</i>	c.4301C>T, p.Ser1434Leu	68	1
	L1200	<i>XPC</i>	c.860T>G, p.Phe287Cys	64	0,5

Legenda: Todas as reações de amplificação por PCR foram efetuadas com 35 ciclos.

IV. Condições de PCR utilizadas para a amplificação das variantes selecionadas através da análise bioinformática, a partir da análise de WES, para o estudo de segregação com a doença na família L56

Output	Gene	Variante amplificada	Temperatura (°C)	MgCl2 (µl)
4	<i>C16ORF96</i>	c.611C>T ;p.Ser204Phe	59,5	0,5
	<i>C20RF70</i>	c.232C>G;p.Pro78Ala	60	0,5
	<i>DCHS1</i>	p.Cys34_Cys35ins	60	0,5
	<i>DUSP12</i>	c.6_10insTTG; p.Leu2_Leu3ins;	60	0,5
	<i>GP1BA</i>	c.775G>A;p. Val259Met	62	0,5
	<i>MAGI1</i>	Gln421_Thr422insGln	56	0,5
	<i>MAML2</i>	c.1821_1829delACAGCAGCA;p.Gln472_Gln474del	63,5	0,5
	<i>MTMR3</i>	c.1933C>T;p.Arg645Trp	60,5	0,5
	<i>NPR2</i>	c.2644G>A;p.Val882Ile	62	1
	<i>OR3A1*</i>	c.392C>T;p.Arg131Gln	69	2
	<i>OR56A4</i>	c.464T>C; p.Q155R	58	0,5
	<i>PROSER2**</i>	c.607G>A; p.Glu203Lys	66	1
	<i>RNF207</i>	c.1615C>T; p.Arg2695Cys	59	0,5
	<i>SMG7</i>	c.1757C>A; p.Pro657His	59	1,5
3	<i>TNIP3</i>	c.854T>G; p.Glu285Ala	60	0,5
	<i>AKRIC3</i>	c.623C>T;p.Ser208Leu	65	1
	<i>ANKRD12</i>	c.297 ^A >G; p.Thr99=	63	1
	<i>ARHGAP22</i>	c.963G>A; p.Asn321=	66	1
	<i>BBOX1</i>	c.639+5_639+6insG	62	0,5
	<i>BRD4</i>	c.3810C>G; p.Glu1270Asp	65	0,5

	<i>CEP250</i>	c.4208A>C; p.Gln1403Pro	65	1
	<i>CLIP2</i>	c.2238G>A; p.Ala746=	65	0,5
	<i>GALNT8</i>	c.988G>A; p.Glu330Lys	64,5	0,5
	<i>KIAA1462</i>	c.1660T>C; p.Thr554Ala	65	1,25
	<i>LGR6</i>	c.1868G>A; p.Arg623Gln	65	1
	<i>MAP4</i>	c.6422G>A; p.Ser2141Phe	65	1
	<i>MAP6***</i>	c.1716A>G; p.Asn572=	64	0,5
	<i>NEK7</i>	c.103C>G; p.Arg35Gly	63	1
	<i>PARD3</i>	c.2540G>A; p.Ala847Val	65	1
	<i>PELP1</i>	c.894C>T; p.Ser298=	64	1
	<i>PFKP</i>	c.339C>T; p.Asn113=	65	1
	<i>PTPN22</i>	c.197-4A>G	63	1
	<i>PTPN23</i>	c.3988G>A; p.Val1330Met	64,5	1,25
	<i>RASL11B</i>	c.577G>A; p.Ala193Thr	65	1,25
	<i>RNF217</i>	c.6+3A>G	65	0,5
	<i>SULT1A1</i>	c.-44C>A	65	0,5
	<i>TM4SF5***</i>	c.179T>C; p.Val60Ala	60,5	1
	<i>TP63</i>	c.-72 -71insAGAG	62	0,75
	<i>WDR37</i>	c.531C>T; p.Ala177=	65	1
	<i>ZNF75A</i>	c.A931G	60	0,5
Análise Bioinformática Anterior	<i>CYB5R2</i>	c.55C>T; p.Pro19Ser	64	0,5
	<i>DNAH3</i>	c.10915C>A; p.Leu3639Ile	59,5	0,5
	<i>OBSCN</i>	c.8365G>C; p.Gly2789Arg	60,5	0,5

Legenda: * Adição de 2 µl de solução *GC rich resolution PCR System (Roche)*; ** Adição de 3% de DMSO; *** Utilização de reagentes *AmpliTaQ Gold® (Thermofisher)*; Todas as reações de amplificação por PCR foram efetuadas com 35 ciclos.

V. Condições de PCR utilizadas para a análise somática da variante p.Arg645Trp do gene *MTMR3*

Gene	Variante amplificada	Temperatura (°C)	MgCl ₂ (ul)	Ciclos	Reagentes
<i>MTMR3</i>	p.Arg645Trp	69,5	1,65	40	<i>AmpliTaQ Gold (Thermofisher Scientific)</i>

VI. Condições de qPCR e amostras utilizadas para as análises de expressão dos genes *MTMR3*, *RAC1* e *RAC1b*

Análise	Gene	Amostras	Concentração dos primers	Concentração da amostra	Kit utilizado	Eficiência da Recta
Análise de expressão	<i>MTMR3</i>	cDNA: L674T, Mucosa B, Mucosa C, Cas638C, A193C, <i>Sw480</i>	5 pmol/ul	15 ng/ul	<i>KapaBiosystems</i>	2,13
	<i>RAC1</i>		3 pmol/ul		<i>Applied Biosystems</i>	2,18
	<i>RAC1b</i>		5 pmol/ul		<i>KapaBiosystems</i>	2,24
	<i>GADPH</i>		5 pmol/ul		<i>KapaBiosystems</i>	2,12

Anexo C: Programas de amplificação por PCR utilizados neste estudo

I. Programa utilizado na amplificação por PCR com os reagentes do *kit BiotaqTM* (Bioline)

Etapa	Temperatura (°C)	Duração	Nº Ciclos
Desnaturação inicial	95	5 minutos	1
Desnaturação	94	50 segundos	30-35
<i>Annealing</i>	Variável	30 segundos	
Extensão	72	50 segundos	
Extensão Final	72	7 minutos	1
Pausa	15	∞	∞

II. Programa utilizado na amplificação por PCR com os reagentes do *kit AmpliTaq Gold[®]* (Thermofisher)

Etapa	Temperatura (°C)	Duração	Nº Ciclos
Desnaturação inicial	95	10 minutos	1
Desnaturação	94	50 segundos	30-35
<i>Annealing</i>	Variável	30 segundos	
Extensão	70	1 minuto	
Extensão Final	70	5 minutos	1
Pausa	15	∞	∞

III. Programa utilizado na reacção de sequenciação para os genes em estudo

Etapa	Temperatura (°C)	Duração	Nº Ciclos
Desnaturação inicial	96	5 minutos	1
Desnaturação	95	10 segundos	25
<i>Annealing</i>	59,5	5 segundos	
Extensão	60	4 minutos	
Pausa	4	∞	∞

IV. Programa utilizado na reacção de transcrição reversa para síntese de cDNA

Etapa	Temperatura (°C)	Duração (min)	Nº de Ciclos
Desnaturação e ligação dos hexâmeros	70	10	1
Pausa	10	∞	
Síntese do cDNA	42	60	
Extensão	70	15	
Pausa	4	∞	∞

- V. Programa utilizado na reacção qPCR para a análise de expressão e *copy number* do gene *FBXO11*

Etapa	Temperatura (°C)	Duração	Ciclos
Incubação	50	2 minutos	-
Ativação da polimerase	95	10 minutos	-
Desnaturação	95	15 segundos	40
Hibridação e extensão	60	1 minutos	

- VI. Programa utilizado na reacção qPCR para a análise de expressão dos genes *MTMR3*, *RAC1* e *RAC1b*.

Etapa	Temperatura (°C)	Duração	Ciclos
Incubação	50	2 minutos	-
Ativação da polimerase	95	10 minutos	-
Desnaturação	95	25 segundos	45
Hibridação e extensão	61	1 minuto e 40 segundos	

Anexo D: Preparação dos reagentes utilizados na electroforese em gel de agarose

1. Preparação do tampão de electroforese:

Para se obter TBE 1x, procedeu-se a uma diluição a partir de TBE 10x (0,89M Tris Borato pH8,3 + 20mM Na₂ EDTA – *National Diagnostics*) para um volume final de 2000ml, em água bidestilada.

2. Preparação do gel de agarose 2% (p/v):

- I. Pesar 5g de agarose (Seaken® LE Agarose – Lonza), para um *erlenmeyer* de 500ml;
- II. Adicionar 250ml de tampão TBE 1x;
- III. Dissolver a solução no microondas;
- IV. Adicionar 12,5µl de brometo de etídio (10mg/ml, *MP biomedical*s) e agitar para homogeneização;
- V. Colocar a solução num molde com 4 pentes e deixar arrefecer até à solidificação do gel.

3. Preparação do gel de agarose 0,8% (p/v)

- I. Pesar 1,2g de agarose (Seaken® LE Agarose – Lonza), para um *erlenmeyer* de 500ml;
- II. Adicionar 150ml de tampão TBE 1x;
- III. Dissolver a solução no microondas;
- IV. Adicionar 7,5µl de brometo de etídio (10mg/ml, *MP biomedical*s) e agitar para homogeneização;
- V. Colocar a solução num molde com 1 pentes e deixar arrefecer até à solidificação do gel.

4. Preparação do *Orange G* 5x e 1x:

4.1. *Orange G* 5x: preparado a partir de 12ml de Ficoll (*Sigma*), 125g de *Orange G* (*Sigma*) e 50ml de água bidestilada, sob agitação. Aliquotar e armazenar a -20°C.

4.2. *Orange G* 1x: diluído a partir de *Orange G* 5x para um volume final de 1ml em água bidestilada. Armazenar a 4°C.

5. Preparação do marcador de peso molecular *GeneRuler 50bp DNA Ladder* (*ThermoFisher Scientific*):

Adicionar 50µl de *GeneRuler 50bp DNA Ladder* (*ThermoFisher Scientific*) a 250µl de *Orange G* 5x e perfazer o volume para 1ml com água bidestilada.

6. Preparação do marcador molecular *Lambda/HindIII DNA Ladder* (*Fermentas*)

Adicionar 0,5µl de marcador *Lambda/HindIII DNA Ladder* (*Fermentas*) a 2,5µl de *Orange G* 5x e 7µl de água bidestilada. Este marcador, por não ser tão utilizado, é preparado de fresco, quando necessário.

Anexo E: Protocolo de precipitação e purificação do DNA Etanol/EDTA/Acetato de sódio – *BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems)* após a reacção de sequenciação

1. Num tubo tipo *eppendorf* de 2ml adicionar 2µl de EDTA (125mM), 2µl de acetato de sódio (3M) e 50µl de etanol absoluto, por cada reacção;
2. Agitar em *vortex* e fazer *spin-down*;
3. Transferir 54µl do sobrenadante para novos tubos tipo *eppendorf* de 1,5ml, adicionar o volume total da reacção de sequenciação e homogeneizar;
4. Incubar durante 15minutos à temperatura ambiente;
5. Centrifugar a 14000rpm durante 30minutos a 4°C;
6. Retirar completamente o sobrenadante com uma micropipeta;
7. Adicionar 100µl de etanol a 70% (v/v) e agitar ligeiramente em *vortex*;
8. Centrifugar a 14000rpm durante 15minutos a 4°C;
9. Retirar completamente o sobrenadante com uma micropipeta;
10. Secar o *pellet* a 37°C num banho-seco durante cerca de 10minutos;
11. Guardar o *pellet* a 4°C.

I. EDTA (125 mM, pH 8)

Dissolver 4,65g de EDTA em água destilada e perfazer o volume para 100 mL.

II. Acetato de sódio (3M, pH 4.5)

Dissolver 24,8g de acetato de sódio em água destilada. Acertar o pH com ácido acético e perfazer para um volume de 100 mL.

III. Etanol 70%

A 70 mL de etanol absoluto, adicionar 30 mL de água destilada e homogeneizar.

Anexo F: I- Constituição do painel multigénico (Trusight Cancer Illumina) de 94 genes associados ao aumento de risco para cancro

Genes incluídos no painel multigénico			
<i>AIP</i>	<i>EGFR</i>	<i>HNFI1A</i>	<i>RB1</i>
<i>ALK</i>	<i>EPCAM</i>	<i>HRAS</i>	<i>RECQL4</i>
<i>APC</i>	<i>ERCC2</i>	<i>KIT</i>	<i>RET</i>
<i>ATM</i>	<i>ERCC3</i>	<i>MAX</i>	<i>RHBDF2</i>
<i>BAP1</i>	<i>ERCC4</i>	<i>MEN1</i>	<i>RUNX1</i>
<i>BLM</i>	<i>ERCC5</i>	<i>MET</i>	<i>SBDS</i>
<i>BMPRI1A</i>	<i>EXT1</i>	<i>MLH1</i>	<i>SDHAF2</i>
<i>BRCA1</i>	<i>EXT2</i>	<i>MSH2</i>	<i>SDHB</i>
<i>BRCA2</i>	<i>EZH2</i>	<i>MSH6</i>	<i>SDHC</i>
<i>BRIP1</i>	<i>FANCA</i>	<i>MUTYH</i>	<i>SDHD</i>
<i>BUB1B</i>	<i>FANCB</i>	<i>NBN</i>	<i>SLX4</i>
<i>CDC73</i>	<i>FANCC</i>	<i>NF1</i>	<i>SMAD4</i>
<i>CDH1</i>	<i>FANCD2</i>	<i>NF2</i>	<i>SMARCB1</i>
<i>CDK4</i>	<i>FANCE</i>	<i>NSD1</i>	<i>STK11</i>
<i>CDKN1C</i>	<i>FANCF</i>	<i>PALB2</i>	<i>SUFU</i>
<i>CDKN2A</i>	<i>FANCG</i>	<i>PHOX2B</i>	<i>TMEM127</i>
<i>CEBPA</i>	<i>FANCI</i>	<i>PMS1</i>	<i>TP53</i>
<i>CEP57</i>	<i>FANCL</i>	<i>PMS2</i>	<i>TSC1</i>
<i>CHEK2</i>	<i>FANCM</i>	<i>PRF1</i>	<i>TSC2</i>
<i>CYLD</i>	<i>FH</i>	<i>PRKARIA</i>	<i>VHL</i>
<i>DDB2</i>	<i>FLCN</i>	<i>PTCH1</i>	<i>WRN</i>
<i>DICER1</i>	<i>GATA2</i>	<i>PTEN</i>	<i>WT1</i>
<i>DIS3L2</i>	<i>GPC3</i>	<i>RAD51C</i>	<i>XPA</i>
		<i>RAD51D</i>	<i>XPC</i>

Anexo G: I-Resultados da análise mutacional efectuada nos 22 exões do gene *FBXO11* para as 24 amostras de DNA (15 de indivíduos index das famílias FCCTX e 3 de familiares afetados e 6 de indivíduos HNPCC e/ou que preenchem os CB, cujos tumores apresentavam ausência de expressão do gene *MSH6*, sem mutação germinal identificada nos genes MMR).

	DNA	Família	Resultados da análise mutacional do gene <i>FBXO11</i>															
			Ex 1/2	Ex 3	Ex 4	Ex 5/6	Ex 7/8	Ex 9/10	Ex 11	EX 12	Ex 13	Ex 14/15	Ex 16	Ex 17	Ex 18	Ex 19	Ex 20	Ex 21/22
Índex	L5	7	Normal	Normal	Normal	Normal	Normal*	Normal	Normal	Normal	Het 6*	Normal*	Normal	Normal	Normal*	Normal*	Het 9	Normal*
	L48	24	Normal	Normal	Normal	Normal	Normal*	Normal	Normal	Normal	Normal*	Normal*	Normal	Normal	Normal*	Normal*	Normal	Normal*
	L108	37	Het1	Normal	Normal	Normal	Normal*	Normal	Normal	Normal	Normal*	Normal*	Normal	Normal	Normal*	Normal*	Normal	Normal*
	L295	55	Normal	Normal	Normal	Normal	Normal*	Normal	Normal	Normal	Normal*	Normal*	Normal	Normal	Normal*	Normal*	Normal	Normal*
	L926		Normal	Normal	Normal	Normal	Normal*	Normal	Normal	Normal	Normal*	Normal*	Normal	Normal	Normal*	Normal*	Normal	Normal*
	L408	56	Normal	Normal	Normal	Normal	Normal*	Normal	Normal	Normal	Normal*	Normal*	Normal	Normal	Normal*	Normal*	Normal	Normal*
	L674		Normal	Normal	Normal	Normal	Normal*	Normal	Het 4	Normal	Normal*	Normal*	Normal	Normal	Normal*	Normal*	Normal	Normal*
	L451	63	Normal	Normal	Normal	Normal	Het 2*	Normal	Normal	Normal	Normal*	Normal*	Normal	Normal	Normal*	Normal*	Normal	Normal*
	L720	63	Normal	Normal	Normal	Normal	Normal*	Normal	Normal	Normal	Normal*	Normal*	Normal	Normal	Normal*	Normal*	Normal	Normal*
	E436	69	Normal	Normal	Normal	Normal	Normal*	Normal	Normal	Normal	Normal*	Normal*	Normal	Normal	Normal*	Normal*	Normal	Normal*
	L467	71	Normal	Normal	Normal	Normal	Normal*	Normal	Normal	Normal	Normal*	Normal*	Normal	Normal	Normal*	Normal*	Normal	Normal*
	L484	84	Normal	Normal	Normal	Normal	Normal*	Normal	Normal	Normal	Normal*	Normal*	Normal	Normal	Normal*	Normal*	Normal	Normal*
	L499	86	Normal	Normal	Normal	Normal	Normal*	Normal	Normal	Normal	Normal*	Normal*	Normal	Normal	Normal*	Normal*	Normal	Normal*
	L701	106	Normal	Normal	Normal	Normal	Normal*	Normal	Normal	Normal	Normal*	Normal*	Normal	Normal	Normal*	Normal*	Normal	Normal*
	L984	148	Normal	Normal	Normal	Normal	Normal*	Normal	Normal	Normal	Normal*	Normal*	Normal	Normal	Normal*	Normal*	Normal	Normal*
	L1038	173	Normal	Normal	Normal	Normal	Normal*	Normal	Normal	Normal	Normal*	Normal*	Normal	Normal	Normal*	Normal*	Normal	Normal*
	L1200	236	Normal	Normal	Normal	Normal	Normal*	Normal	Normal	Normal	Normal*	Normal*	Normal	Normal	Normal*	Normal*	Normal	Normal*
	L1121	P2209	Normal	Normal	Normal	Normal	Normal*	Normal	Normal	Normal	Normal*	Normal*	Normal	Het 7	Normal*	Normal*	Normal	Normal*
	L1152	L316	Normal	Normal	Normal	Normal	Normal	Het 3	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal
	L1484	L302	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Het 8	Normal	Normal	Normal	Normal
	L1624	P3057	Normal	Normal	Normal	Normal	Normal	Normal	Het 5	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal
	L1635	P3369	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal
	L1846	P3821	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal
	L1855	P3660	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Het 8	Normal	Normal	Normal	Normal

Legenda: Cada individuo foi identificado como Het quando se detectou alguma alteração e normal quando não; Os indivíduos a cinzento correspondem aos 6 indivíduos com tumores com ausência de expressão do gene *MSH6*; Os resultados * foram obtidos da análise mutacional iniciada anteriormente ao projecto; Het 1: p.Thr126Ser; Het 2: c.789+37C>T; Het 3: c.1398+13_1398+16delATTT; Het4: c.1399-107_1399-106insAA; Het5: c.1578A>G; Het6: c.1451-11G >T; Het 7: c.2083+25_2083+26insT; Het 8: c.2227+88T>G; Het 9: c.2555+53T>C;

